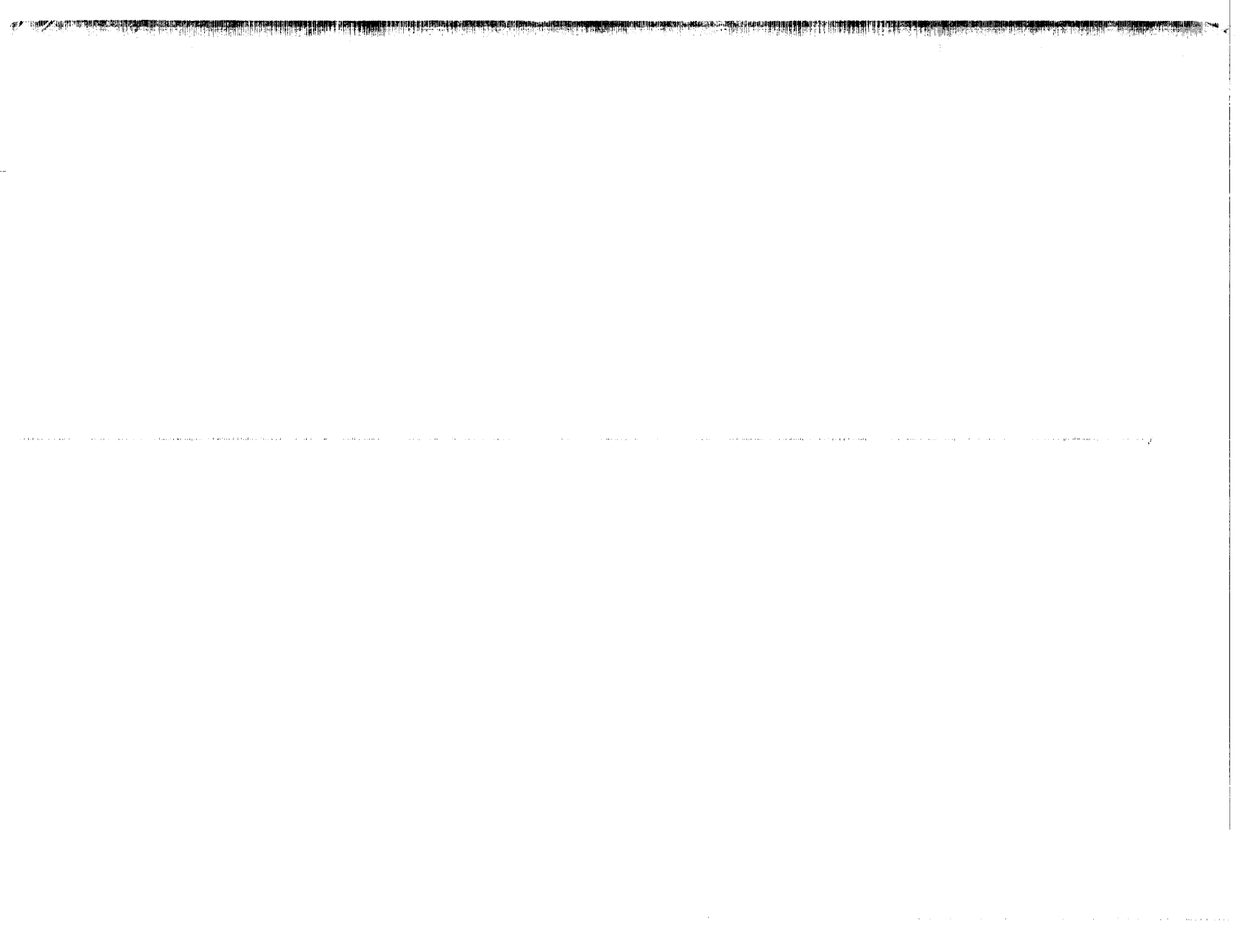


*NASA Conference Publication 3109, Vol. 1*

# Technology 2000

Uncl. is

01/92 0002480



*NASA Conference Publication 3109, Vol. 1*

# **Technology 2000**

Proceedings of a conference held at  
the Washington Hilton Hotel and Towers  
Washington, D.C.  
November 27–28, 1990



National Aeronautics and  
Space Administration

Office of Management

Scientific and Technical  
Information Division

**1991**





## **TECHNOLOGY 2000**

### **SYMPOSIA PROCEEDINGS**

**Presented November 27 & 28, 1990  
Washington, D.C.**

TECHNOLOGY 2000 was the first major industrial conference and exposition spotlighting NASA technology and technology transfer. Its purpose was, and continues to be, to increase awareness of existing NASA-developed technologies that are available for immediate use in the development of new products and processes, and to lay the groundwork for the effective utilization of emerging technologies.

In addition to an exhibit showcasing the products and technologies available for sale or license, TECHNOLOGY 2000 featured 16 concurrent technical sessions in which 97 papers were presented. The program included symposia on Computer Technology and Software Engineering; Human Factors Engineering and Life Sciences; Information and Data Management; Materials Science; Manufacturing and Fabrication Technology; Power, Energy and Control Systems; Robotics; Sensors and Measurement Technology; Artificial Intelligence; Environmental Technology; Optics and Communications; and, Superconductivity.

We are pleased to provide the proceedings from these sessions. They have been published in two volumes. Volume One contains the papers presented Tuesday, November 27th; Volume Two contains the papers presented Wednesday, November 28th.

This is Volume One. It consists of 51 papers. The papers appear in the order in which they were presented at TECHNOLOGY 2000.

For information regarding additional copies, please contact:

**ASSOCIATED BUSINESS PUBLICATIONS  
41 EAST 42ND STREET  
NEW YORK, NY 10017-5391  
(212) 490-3999**



## TABLE OF CONTENTS

### TUESDAY NOVEMBER 27TH

#### SESSION A: COMPUTER TECHNOLOGY AND SOFTWARE ENGINEERING (PART 1)

The Virtual Environment Display System. . . . .	3
Virtual Acoustics Displays. . . . .	10
FAST: A Multi-Processed Environment For Visualization Of Computational Fluid Dynamics. . . . .	21
Hypercube Technology. . . . .	31
The Hyperswitch Network. . . . .	42
Biological Neural Networks As Model Systems For Designing Future Parallel Processing Computers. . . . .	53

#### SESSION B: HUMAN FACTORS ENGINEERING AND LIFE SCIENCES (PART 1)

Biomedical Applications Of NASA Technology. . . . .	63
Direction-Discriminating Hearing Aid System. . . . .	66
X-Ray Imaging Microscope For Cancer Research. . . . .	73
Mechanical Response Tissue Analyzer For Estimating Bone Strength. . . . .	83
Adaptations Of NASA Technology For The Optimization Of Orthopedic Knee Implants. . . . .	89

#### SESSION C: INFORMATION AND DATA MANAGEMENT

Optical Storage Device. . . . .	101
Rewritable Optical Disk Recorder Development. . . . .	109
Monitoring And Analysis Of Data From Complex Systems. . . . .	119
High Data Rate Systems For The Future. . . . .	127
DAVID: The Distribute Access View Integrated Database. . . . .	137
Three-Dimensional Perspective Visualization. . . . .	144

#### SESSION D: MATERIALS SCIENCE (PART 1)

High-Performance Polymer Development. . . . .	149
Industrial Applications Of Graphite Fluoride Fibers. . . . .	156
Unique Applications Of Fluorepoxy Materials. . . . .	165
Dual Beam Process Diamond-Like Films For Industrial Applications. . . . .	173
Plasma-Polymerized Coating For Polycarbonate: Single-Layer, Abrasion-Resistant, And Anti-Reflective. . . . .	184
The PM200 Lubrication System. . . . .	189

#### SESSION E: MANUFACTURING AND FABRICATION TECHNOLOGY

Robotics In Space-Age Manufacturing. . . . .	199
Variable Polarity Plasma Arc Welding. . . . .	208
High-Pressure Water Jet Cutting And Stripping. . . . .	214
Cost-Efficient Manufacturing Of Composite Structures. . . . .	221
Rapid Induction Bonding Of Composites, Plastics, And Metals. . . . .	231
Reliability And Risk Assessment Of Structures. . . . .	241
A Semi-Automated Process For The Production Of Custom-Made Shoes. . . . .	249
Lightweight, Fire-Retardant, Crashworthy Aircraft Seat Cushioning. . . . .	254

PRECEDING PAGE BLANK NOT FILMED

## TABLE OF CONTENTS

### TUESDAY NOVEMBER 27TH (cont'd)

#### SESSION F: POWER, ENERGY, AND CONTROL SYSTEMS

Civil Air Transport: A Fresh Look At Power-By-Wire And Fly-By-Light. . . . .	263
The Free-Piston Stirling Engine - From Space Technology To Terrestrial Applications. . . . .	268
Solar-Powered Stirling Cycle Electricity Generator. . . . .	269
Four Quadrant Control Of Induction Motors. . . . .	279
High Power Bi-Polar Lead-Acid Batteries. . . . .	287
Scientific CCD Technology At JPL. . . . .	294
Advanced Thermal Technology For Commercial Applications. . . . .	301

#### SESSION G: ROBOTICS

ROBOSIM: A Simulator For Robotic Systems. . . . .	309
Control Systems Software, Simulation, And Robotic Applications. . . . .	315
Telerobotic Electronic Materials Processing Equipment. . . . .	322
Advanced Mechanism For Robotics. . . . .	328
The Flight Telerobotic Servicer And Technology Transfer. . . . .	336
FARMS: The Flexible Agricultural Robotics Manipulator System. . . . .	344
Diverse Applications Of Advanced Man-Telerobot Interfaces. . . . .	350

#### SESSION H: SENSORS AND MEASUREMENT TECHNOLOGY (PART 1)

Urodynamic Pressure Sensor. . . . .	363
Electron Tunnel Sensor Technology. . . . .	370
Practical Approaches For The Application Of Resistance-Type Strain Gages On High-Temperature Composites. . . . .	382
AI Mass Spectrometer For Shuttle Health Monitoring. . . . .	392
Instrumentation For Optical Ocean Remote Sensing. . . . .	400
Monitoring And Control Of Atmosphere In A Closed Environment. . . . .	413

**SESSION A - COMPUTER TECHNOLOGY AND SOFTWARE ENGINEERING  
(PART 1)**

**Tuesday November 27, 1990**

- **The Virtual Environment Display System**
- **Virtual Acoustics Displays**
- **FAST: A Multi-Processed Environment For Visualization Of Computational Fluid Dynamics**
- **Hypercube Technology**
- **The Hyperswitch Communications Network**
- **Biological Neural Networks As Model Systems For Designing Future Parallel Processing Computers**



## THE VIRTUAL ENVIRONMENT DISPLAY SYSTEM

Michael W. McGreevy, Ph.D.

Principal Investigator  
NASA Ames Research Center  
Moffett Field, CA 94035

### ABSTRACT

Virtual Environment technology is a display and control technology that can surround a person in an interactive computer-generated or computer-mediated virtual environment. It has evolved at NASA Ames Research Center since 1984 to serve NASA's missions and goals. The exciting potential of this technology, sometimes called Virtual Reality, Artificial Reality, or Cyberspace, has been recognized recently by the popular media, industry, academia, and government organizations. Considerable research and development will be necessary to bring it to fruition.

### INTRODUCTION

Virtual Environment technology provides powerful user interface techniques for use in NASA's mission-oriented applications, including computational fluid dynamics, planetary data visualization, in-space telepresence, and space station telerobotics. In a broader context, several years of international interest in NASA's Virtual Environment technology and related user interfaces has resulted in massive media coverage (e.g. References 1-8), and has encouraged several American companies to develop related commercial products<sup>9</sup>. Using head-tracked head-mounted displays, a Virtual Environment system provides the user with a vivid experience of three-dimensional space. It can be used with computer graphics systems as a personal simulator, surrounding the user with a virtual interactive environment. Alternatively, it can utilize head-slaved cameras, and other sensors, to provide telepresence. In either case, an instrumented glove or similar device may be used to detect hand shape and position, so as to enable the user to manipulate objects in computer generated or remote environments.

The human desire to create virtual environments, synthetic experiences, and artificial realities is an ancient dream of emperors, has a rich recent history with many talented contributors, is a central theme of technology, and is a fundamental aspect of human thinking.<sup>10</sup> The author's original inspiration to actually create virtual environments derived, appropriately enough, from popular media coverage of NASA's unmanned Surveyor missions to the moon in the mid-1960's. In July 1966, Life magazine<sup>11</sup> carried photographs of one of NASA's first virtual environment displays, spherical mosaicked panoramas that allowed scientists to "see what Surveyor's swiveling camera saw." Mimicking the technique, the effects were dramatic.

Today, NASA's Virtual Environment technology allows scientists to explore planetary environments using head-tracked head-mounted displays and computer generated imagery. The computer graphics technology that enables this was originally developed by such pioneers as Ivan Sutherland. His head-mounted display project<sup>12</sup> in the mid-1960's served not only as further inspiration, but it also led to the development of the foundations of computer graphics techniques and systems. While multi-million dollar military systems have used head-mounted displays in the years since Sutherland's work, the notion of a personal virtual environment system as a general purpose user-computer interface was temporarily neglected. In 1984, NASA was able to capitalize on emerging technologies in order to make a dramatic cost reduction in the essential display system, and thus to create the Virtual Environment Workstation<sup>13</sup>. The exciting potential of Virtual Environment technology has now inspired many people to carry it forward to greater capability.

### DESCRIPTION OF THE TECHNOLOGY

A virtual workstation is a general purpose human-information interface device for creating virtual environments that can be optimized for use in specific applications. It is a generalization of today's desktop workstation (with its display screen, keyboard, and mouse). Such workstations provide the mechanism whereby humans interactively process information. The virtual workstation greatly increases the ways in which information can be handled, and can provide a far more intuitive means of interaction with

information. The entire virtual three-dimensional world of the user becomes the display and controls. Every object or data point in that virtual world can be an interactive device. Each can be manipulated by touch, by gesture, or by voice. This vastly broadens the range of expression between the user and the information system far beyond the mere typing of keys or the two dimensional movements of a mouse<sup>14</sup>.

A virtual workstation consists of body-ported devices and the computer/video system hardware and software that supports and integrates them. The user wears a viewer consisting of a video display screen (with its electronics) and wide-angle magnifying lenses. The viewer is supported by a box of electronics for power supply, signal conversions, and functional adjustments and accessories. The viewer is mounted so as to replace the visual field of the user. (A few versions allow the imagery to be combined with the real visual environment of the user.) The head-mounted unit also carries a microphone, earphones, and a head-position and orientation sensor. Typically, the hands are fitted with shape, position, and orientation sensors. Ultimately, the entire body could be tracked.

A host computer integrates the peripheral units and coordinates their interaction. A computer graphics system generates three-dimensional objects for display in the viewer. The imagery is altered in accordance with the head-movements of the user to provide a stable visual environment. Remote video cameras can be slaved to the motions of the user's head, providing a remote source of imagery for visual telepresence. Hand gestures are also interpreted as desired for alteration of the appearance of the scene, behavior of objects, or other interactions with the environment. Telerobotic systems can be controlled by the motions of the user's hands, providing remote manipulation. Peripheral systems convert voice commands to computer commands, convert computer output to synthetic speech or other sounds, and convert body movements to computer input. Video recorders can record what the user sees and hears for non-interactive replay later. Force feedback can be used to allow remote or synthetic objects to feel more solid and real.

The software consists of many reusable modules of computer commands, and various special purpose programs which use these modules and also contain considerable amounts of unique computer code. Digital models of various task environments, such as the Space Shuttle, Space Station, and planetary terrain, are created and stored in databases. They are displayed and manipulated by the software programs under control of the user wearing the virtual workstation.

The virtual workstation will find widespread use in applications involving highly spatial information. These include planetary exploration, telerobotics, computer-aided-design in general and architecture in particular, and scientific visualization. It will also impact applications which can be cast into a spatial context (e.g. three dimensional libraries of text or other "dataspaces"). Further, it may be feasible to create virtual, dynamic three-dimensional worlds from inherently non-spatial data so as to ease interpretation and interaction. Virtual environment systems will greatly influence art, entertainment, education, medicine, and many other fields.

## **MISSION-ORIENTED RESEARCH UTILIZING THE VIRTUAL WORKSTATION**

The creation and generic development of the NASA Virtual Environment Workstation was accomplished under the base research and technology program in Space Human Factors of the NASA Office of Aerospace and Exploration Technology . After several years of generic development and demonstrations, the focus is now on refining the device and its style of interaction for specific NASA applications.

A cooperative project has been established with the Applied Research Office of the Numerical Aerodynamic Simulation (NAS) Systems Division at NASA Ames to develop the Virtual Workstation for NAS applications. Computational fluid dynamicists will utilize the Virtual Workstation in their analyses, while human factors researchers will investigate improved interfaces for the specific visualizations important to CFD.

The objective of the Visualization for Planetary Exploration program is to conduct research and development of crew interfaces for terrain exploration systems. The approach has been: 1) to review mission operational experience, mission constraints and opportunities, and the state-of-the-art in exploration technology, 2) to investigate user behaviors and requirements, and, 3) to enlist interdisciplinary expertise to develop, implement, demonstrate, and evaluate advanced crew interfaces. Field studies have been conducted in desert terrain of interest to planetary geologists and mission planners in order to understand operations



central to planetary surface exploration<sup>15</sup>. In addition, a planetary terrain visualization testbed is under development to support focussed user interface research. This powerful computer system provides dynamic interaction with planetary terrain data, currently Mars and Earth, but is easily applicable to Venus, Earth's moon, or other planetary bodies.

Planetary geoscientists have expressed a strong scientific interest in the development of Virtual Environment technology for planetary surface exploration using telepresence<sup>16</sup>. A leading concept is to extend the reach of human exploration beyond the perimeter of manned EVA by use of telepresence from central manned bases to numerous unmanned rovers<sup>17</sup>. Geoscientists with decades of experience in field work believe that human presence contributes a tightly coupled and essential interplay of cognitive, perceptual, manipulative, and locomotor skills that can complement automated rover operations. By using telepresence, these unique human skills may be applied over a greater area at an earlier date.

The Virtual Environment technology is of interest to the Office of Space Flight for possible use on Space Station Freedom as it evolves. Under the Advanced Development Program, a study is underway at Marquette University and Astronautics Corporation of America to determine the design accommodations, also known as "hooks and scars", that may be required to support eventual use of the Virtual Environment technology onboard Space Station Freedom. Later phases of the study will determine the cost versus benefit trade-offs, and operationally realistic prototypes will be developed.

## **RESEARCH AND DEVELOPMENT ISSUES**

The key research issues in Virtual Environment technology involve the human factors of body-ported displays and controls, the specific needs of the users in the targeted applications, and the computer sciences of system architecture, data management, and computer graphics. Finding the most appropriate engineering tradeoffs for a specific application is one of the most challenging problems. This requires a thorough knowledge of the visual, proprioceptive, vestibular, tactile, and auditory interfaces, including both the human sensory channels and the system capabilities available or required to interact with these channels. Cutting across this view of the problem are the needs, constraints, and opportunities of the application, the user, and the mission.

The special requirements and constraints of this work center around the comprehensive view of the user as an intelligent actor and multi-channel information processor in a completely programmable information environment surrounding the user. Some have referred to this kind of interactive environment as an artificial reality, which certainly implies a massive computational demand. This places a significant burden on the hardware and software of the system.

Most existing user interface equipment is designed for a far less ambitious kind of interaction and a far more constrained model of the user. Thus, prototypes must be built up from systems that were not necessarily intended to function together or for our purposes. In addition, the information bandwidth of the human operator is far in excess of current technology, particularly with respect to real-time rendering of realistic, interactive objects and environments. This places significant constraints on the application of virtual environment interfaces. The challenge is to integrate and apply the best possible system within the many constraints, while also working to advance the technology so as to eliminate the most restrictive constraints.

### **Current impediments to commercial applications**

Before development of Virtual Environment (VE) based commercial applications is reasonable, several basic improvements in the hardware and software are required. The necessary improvements involve enhancements to the functionality and integration of the system components, and providing for adequate complexity of the interactive virtual environment. Given sufficient technical capability, the utility of VEs for real applications must be evaluated in detail. The impediments include limited resolution viewers, lack of available digital models, poor system integration and standardization, and the lack of careful attention to the specific applications.

The visual resolution of available viewers is too low and/or the cost is too high. There have been no mass produced VE viewers, and those that are available are crude. As a result, VE systems suffer from limited

applicability to real-world problems. A mass produced viewer is needed that has high resolution, wide field of view, and low cost.

There are no readily accessible collections of interactive digital models to populate virtual environments. Instead, each group must generate these models from scratch. This results in ad hoc solutions, poor commonality, redundant effort, and limited extensibility. A comprehensive effort is needed to ensure effective design, generation, and distribution of digital models for VEs.

Comprehensive and integrated hardware and software tools for creating VEs are not available. As a result, many potential contributors cannot yet become involved in VE research and development. Currently, the many components of a VE system must be pieced together from existing technology, resulting in poorly integrated systems, ad hoc architectures, and poor commonality. Most current networks, computer graphics systems, processors, and peripherals were not specifically designed for use in VEs, and the "wrong" tradeoffs have been sometimes been made. This results in fundamental limitations in the architecture of current (ad hoc) VE systems. New approaches must be explored and optimal VE architectures must be identified and implemented. A collective effort should promote standardization, while retaining extensibility and the ability to integrate proprietary tools.

In parallel to the technical impediments outlined above, there has not yet been sufficient focussed effort by industry to determine the potential of VEs in commercial applications. As a result, there is limited recognition (outside of Japan) of the potential of this technology. American industry must investigate commercial applications of Virtual Environment technology.

#### **Research and development to advance commercial applications**

Research and development to advance commercial applications of virtual environment technology should concentrate on four major areas: 1) devices, 2) modelling, 3) interactivity, and 4) user requirements.

VE devices are the hardware and software that enable the user to experience and interact with a virtual environment. VE devices includes all of the physical hardware and all of the general purpose system software. Software unique to an application is also included in this component, but is likely to be proprietary. The objectives of work on VE devices should be to create and develop integrated, modular, extensible hardware and software tools for building VEs; to develop tools for VEs that are capable of benefitting high payoff applications; and to make a common set of tools available. And it is important to stress the importance of supporting real-time interactions in Virtual Environments.

Modelling is the acquisition and generation of digital models of environments and objects, along with their attributes and behaviors, which are the contents of a VE, as well as their archiving, distribution, manipulation, and modification. This includes all of the hardware and software that are specific to modelling. The overall objective in modelling work for VE should be to develop tools and techniques that provide for the effective design, capture or creation, and distribution of interactive digital models of objects and environments for VEs.

Interactivity refers to the style or "look and feel" of the VE, the metaphors<sup>18</sup> by which information is organized, the behaviors of objects and the "physics" of the environment as they appear to the user, and the collection of ways that the user may interact with virtual objects and the virtual environment. Research objectives to support commercial applications should include optimization of the effectiveness of the user in accessing, understanding, and acting upon complex information in Virtual Environments. And it is obviously important to support ease of use, robustness, user friendliness, and a minimal need for manuals. The nature of the appropriate "look and feel", metaphors, behaviors, and "physics" are all open questions, and depend upon the applications.

User requirements are a set of specifications based on user needs, problems, constraints, and opportunities that guide the design and implementation of tools intended to help users. Work done in this area will have a significant impact on the design of application hardware and software, of course, but will also drive the interactivity paradigms. Objectives in the area of user requirements should be: identify the needs and problems of users that might be addressed and solved by VE technology; conduct field studies of users in selected candidate applications; develop application testbeds that address real user requirements; and provide

feedback to implementers of VE devices, especially viewers, trackers, and other devices that directly impact on the user's working environment.

## CONCLUSION

Virtual Environment technology has tremendous potential, but considerable research and development is necessary to bring it to fruition. However, the media hype might lead one to believe that the technology is nothing more than "goggles and gloves." This emphasis is understandable given that this hardware is the most obviously new physical part of virtual environment interfaces. The real revolution, once there are lots of goggles and gloves around, will be in the generation of, and interaction with, computer-generated digital worlds, and the ability to be telepresent in remote locations via computer mediation. This will influence NASA's missions and goals, for example, by allowing everyone to personally explore digital models of the canyons of Mars, or by enabling astronauts to do routine maintenance while safely inside the space station. It will also greatly influence art, entertainment, education, medicine, and many other fields. The Virtual Environment team at NASA is proud to have contributed to this advancement in human-centered technology.

## TEAM MEMBERS

The Virtual Environment Research program has been made possible by the combined efforts and dedication of a diverse team of researchers, each of whom has made unique and substantial contributions. The group is devoted to research and development of the concepts, implementation, and applications of virtual environment technology that serve NASA's missions and goals.

### NASA Ames Research Center

Dr. Michael McGreevy is the originator of "Virtual Environment" user interface research at NASA. He is a research scientist, and is Principal Engineer of the Human Interface Research Branch. He leads the Visualization for Planetary Exploration program, the Telepresence User Science Requirements project, and the Space Station Advanced Development program in Virtual Environment displays. Since 1981 he has conducted research in spatial information transfer at NASA. In 1984, he created the Virtual Visual Environment Display (VIVED) program at Ames and designed the VIVED system with the help of Humphries, Eriskin, and Deardon. By early 1985, he designed and developed NASA's first Virtual Environment Workstation, a personal simulator and telepresence device consisting of a host computer, an interactive computer graphics system, video imaging technology, and the VIVED system. Between July 1985 and August 1987, McGreevy was temporarily assigned to NASA Headquarters in Washington D.C, as Program Manager for NASA's research programs in Aerospace Human Factors and Aeronautical Computer Science. For his research management activities in Washington, he won NASA's Special Achievement Award. Since the beginning of the Virtual Environment project in 1984, he has actively promoted technology transfer within NASA, and to other government organizations, industry, and academia.

Dr. Stephen R. Ellis is a research psychologist with a long history of research in spatial instruments. In 1981, he and McGreevy began a program of research in spatial information transfer, emphasizing the interpretation of perspective displays. He is currently developing research projects utilizing the Virtual Workstation for part-task experiments, including studies in manual tracking and stereo vision. He is also an adjunct professor in the Optometry Department at UC Berkeley.

Dr. Beth Wenzel contributed the concept of three-dimensional sound and auditory symbology to the Virtual Workstation project. Together with university scientists, she is investigating the science of 3D sound and applying her findings to the project.

Dr. Jim Larimer joined the project in 1990 and is currently working to upgrade the viewer technology and plans to develop additional refinements based on parameters of human vision.

## **Research Institute for Advanced Computer Science (RIACS)**

Dr. Lew Hitchner is a computer scientist and computational geographer, and is responsible for software implementation of a user interface testbed to support research in visualization for planetary exploration. He was hired by RIACS, an institute of the Universities Space Research Association, in October 1989 specifically to work under contract to NASA on the Visualization for Planetary Exploration program. Prior to this, he was a professor of computer graphics at UC Santa Cruz. Hitchner earned his Ph.D. in Computer Science at University of Utah.

### **Sterling Federal Systems**

Amy Wu assisted McGreevy in programming NASA's first Virtual Workstation, and is currently lab systems manager and applications programmer for the Visualization for Planetary Exploration program.

Jim Humphries, Design Engineer and Hardware Lab Manager, co-designed (with Saim Eriskin) the electronics for the original VIVED system and for the subsequent refinements of the original design, including the first boom-mounted system.

Joe Deardon, Senior Technician, has assisted Jim Humphries and Saim Eriskin in implementation of several generations of the VIVED system since the beginning of the project.

Rick Jacoby has been a research programmer on the program since 1988. He programmed an interactive telerobotics demonstration on the Virtual Workstation and is currently enhancing it for use as a research tool.

Phil Stone is a programmer working for Beth Wenzel who has been with the project since 1988.

### **UC Berkeley**

Romy Bauer is a graduate student in computer science. After receiving a BS in computer science and a BS in geography at UC Berkeley, she worked full time for a year (1989-1990) developing terrain analysis and visualization techniques for the VPE project. She is currently a half-time research assistant on the project.

### **Other Key Technical Contributors**

Eric Howlett of Pop-Optix Labs designed and built the wide angle stereo optics.

Shigeru Morokawa, R&D Manager of the Technical Research Lab, Citizen Watch Company, designed, custom-built, and specially modified many of the liquid crystal display systems needed for the program.

Dr. Xin Feng, Assistant Professor, Dept. of Electrical, Computer, and Biomedical Engineering, Marquette University, Milwaukee, WI, is Principal University Investigator on the Space Station Virtual Environment technology project.

Richard Foster, Manager, Space Systems Dept., Astronautics Corp. of America, Madison, WI, has over 25 years of systems engineering experience and is Principal Industry Investigator on the Space Station VE technology project.

### **Ex-Members of the Team**

Saim Eriskin (Sterling) served as Senior Engineer and Electronics Designer from the first day of the project in 1984 until October 1990. He co-designed (with Jim Humphries) the electronics for the original VIVED system and for the subsequent refinements of the original design, including the first boom-mounted system.

Scott Fisher (NASA) was hired in 1985 to join the growing project team and left NASA in early 1990. He brought the VPL Research Dataglove to the Virtual Workstation, and provided a creative outlook and familiarity with innovative human interface projects and researchers. His experience includes association with related activities at Atari Research and MIT's Media Lab. Mr. Fisher served as lab manager and point of contact at Ames during fiscal years 1986 and 1987.

Jim Tanner and Dave Kaiser (Sterling) were Project Managers during fiscal years 1986 and 1987.

Doug Kerr (Sterling) designed software libraries and database tools, and provided expertise in systems and applications software which were critical to the success of the project between 1985 and 1987.

Warren Robinett (NASA) contributed his extensive software skills and demonstration software 1985-1987.

Stephen Bryson (Sterling) was a programmer on the Virtual Workstation program between 1988 and 1990. He is currently working with the Numerical Aerodynamic Simulation staff on a spinoff virtual environment project.

David Koblas (RIACS) was a research programmer on the VPE program during the summer of 1990. He developed significant application software refinements including improved interactive capabilities.

Mark Bolas (Fake Space Labs) designed and built a motor-driven camera platform in 1989.

Other ex-members include: Cordell Ratzlaff, Ian McDowell, Dr. Bob Brown, Ken Uhland, Steffan Jeffers, Clay Coler, Al Duncan, and a long list of students and part-time programmers.

## ACKNOWLEDGEMENTS

This work is conducted within NASA's leading organization for human factors research, the Aerospace Human Factors Research Division, at NASA Ames Research Center, Moffett Field, California, in the heart of Silicon Valley. The Virtual Environment R&D program continues to be funded by the Human Factors Research Program of NASA's Office of Aerospace and Exploration Technology (OAET). The Office of Space Station's Advanced Development Program also provides major support. Funding for telepresence field studies came from the Office of Exploration.

## REFERENCES

1. Rogers, M.: "Now, 'Artificial reality' ," Newsweek, vol. 111, no. 6, pp. 56-57, 1987.
2. Meigs, J. B.: "3-D TV comes home," Popular Mechanics, vol. 164, no. 8, pp. 67-69, 88, 90, August 1987.
3. Anon.: "NASA's Virtual Workstation: Using computers to alter reality," NASA Tech Briefs, vol. 12, no. 7, July/Aug 1988.
4. Jenish, D'Arcy, "Re-creating reality, " Macleans, vol. 103, no. 23, pp 56-57, June 4, 1990.
5. Daviss, Bennett: "Grand illusions," Discover, vol. 11, no. 6, pp. 36-41, June 1990.
6. Anon.: " 'Cyberspace': Programmerte Paradiese," Der Spiegel, vol. 44, no. 34, pp. 138-143, August 20, 1990.
7. Stark, D. E.: "Journey through cyberspace: The new frontier of virtual reality," High Technology Careers Magazine, vol. 7, no. 4, pp. 6, 9, August/September 1990.
8. Ditlea, S.: "Computerized tour guides," Omni, vol. 13, no. 1, pg. 26, October 1990.
9. Anon.: "Computerized reality comes of age," NASA Tech Briefs, vol. 14, no. 8, pp.10-12, August 1990.
10. McGreevy, M. W.: "Personal simulators and planetary exploration," plenary speech, CHI '89, Association for Computing Machinery conference on Computer-Human Interaction, May 1989. (transcript available)
11. Anon.: "From Surveyor: The stark and airless beauty of the moon," pp. 62-67, Life, vol. 61, no. 1, July 1, 1966.
12. Sutherland, Ivan E.: "A head-mounted three-dimensional display," Proceedings of the Fall Joint Computer Conference, pp. 757-764, 1968.
13. McGreevy, M.W.: NASA Ames Virtual Environment Display: Applications and requirements, internal technical document, Aerospace Human Factors Research Division, NASA Ames Research Center, September 1984.
14. McGreevy, M. W.: "Personal simulators and divergence from realism, " plenary speech, ACM SIGGRAPH Symposium on User Interface Software, October 1988. (partial transcript available)
15. McGreevy, M. W., Stoker, C. R.: "Telepresence for planetary exploration," In Proceedings of the SPIE 1990 Conference on Cooperative Intelligent Robotics in Space (in press).
16. Taylor, G. Jeffrey, and Spudis, Paul D.: "A teleoperated robotic field geologist," In Proceedings of Space '90: Engineering, Construction, and Operations in Space (in press).
17. NASA: Report of the 90-Day Study of Human Exploration of the Moon and Mars, November 1989.
18. McGreevy, M. W.: "The exploration metaphor," invited technical presentation, Engineering Foundation Conference on Human-Machine Interfaces for Teleoperators and Virtual Environments, March 1990. (transcript available)

N 9 1 - 2 3 0 2 3

## VIRTUAL ACOUSTICS DISPLAYS

Elizabeth M. Wenzel  
NASA-Ames Research Center, MS 262-2  
Moffett Field, CA 94035

Scott S. Fisher  
11571 Buena Vista Drive  
Los Altos Hills, CA 94022

Philip K. Stone  
Sterling Software  
NASA-Ames Research Center, MS 262-6  
Moffett Field, CA 94035

Scott H. Foster  
Crystal River Engineering  
12350 Wards Ferry Road  
Groveland, CA 95321

### Abstract

*This paper describes the real time acoustic display capabilities developed for the Virtual Environment Workstation (VIEW) Project at NASA-Ames Research Center. The acoustic display is capable of generating localized acoustic cues in real time over headphones. An auditory symbology, a related collection of representational auditory "objects" or "icons," can be designed using ACE, the Auditory Cue Editor, which links both discrete and continuously-varying acoustic parameters with information or events in the display. During a given display scenario, the symbology can be dynamically co-ordinated in real time with three-dimensional visual objects, speech, and gestural displays. The types of displays feasible with the system range from simple warnings and alarms to the acoustic representation of multidimensional data or events.*

### Introduction

Recent years have seen many advances in computing technology with the associated requirement that users manage and interpret increasingly complex systems of information. As a result, an increasing amount of applied research has been devoted to a type of reconfigurable interface called the virtual display. Some of the earliest work in this area was done by Sutherland [30] at the University of Utah using binocular head-mounted displays. Sutherland characterized the goal of virtual interface research, stating that, "The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real." As the technology has advanced, virtual displays have gone beyond the flat CRT screen, assuming a three-dimensional spatial organization which, it is hoped, provides a richer and more natural means of accessing and manipulating information. A few projects have taken the spatial metaphor to its limit by directly involving the operator in the data environment [5], [19], [21]. Thus, the kind of "artificial reality" once relegated solely to the specialized world of the cockpit simulator is now being seen as a next step in interface development for all types of advanced computing applications [20].

©1990-The Institute of Electrical and Electronics Engineers, Inc. Reprinted, with permission, from *Proceeding of Visualization '90*, San Francisco, California, October 23-26, 1990.

## Auditory Icons & Symbolologies

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel, particularly when the visual channel is overloaded and visual cues are degraded or absent [12], [13], [27]. Most recently, attention has been devoted to the use of non-speech audio as an interface medium [1], [2], [8], [23]. Auditory signals are detected more quickly than visual signals and tend to produce an alerting or orienting response. Consequently, non-speech audio has been most frequently used in simple alarm or warning systems, as in aircraft cockpits or the siren of an ambulance. Another advantage of audition is that it is primarily a temporal sense and we are extremely sensitive to changes in an acoustic signal over time. This feature tends to bring any such acoustical event to our attention and conversely, allows us to relegate sustained or uninformative sounds to the background. Thus audio is particularly suited to monitoring changes over time, for example when your car engine suddenly begins to malfunction. Non-speech signals have the potential to provide an even richer display medium if they are carefully designed with human perceptual abilities in mind. Just as a movie with sound is much more compelling and informationally-rich than a silent film, so could a computer interface be enhanced by an appropriate "sound track" to the task at hand. If used properly, sound need not be distracting or cacophonous or merely uninformative. Principles of design for auditory icons and symbolologies can be gleaned from the fields of music, psychoacoustics, and psychological studies of the acoustical determinants of perceptual organization. For example, one can think of the audible world as being composed of a collection of acoustic "objects." Various acoustic features, such as timbre, intensity, and temporal rhythm, specify the identities of the objects and perhaps convey meaning about discrete events or ongoing actions in the world and their relationships to one another. One could systematically manipulate these features and create an auditory symbology which operates on a continuum from "literal" everyday sounds, such as the clunk of mail in your mailbox (e.g., Gaver's 'Sonic Finder' [23]), to a completely abstract mapping of statistical data into sound parameters [4], [28].

Such a display could be further enhanced by taking advantage of the auditory system's ability to segregate, monitor, and switch attention among simultaneous sources of sound. One of the most important determinants of acoustic segregation is an object's location in space.

A true three-dimensional auditory display could potentially improve information transfer by combining directional and iconic information in a quite naturalistic representation of dynamic objects in the interface. Borrowing a term from Gaver [23], an obvious aspect of "everyday listening" is the fact that we live and listen in a three-dimensional world. Indeed, a primary advantage of the auditory system is that it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. This feature would be especially useful in an application that is inherently spatial, such as an air traffic control display for the tower or cockpit, or even in a two-dimensional interface which has adopted a spatial organization, such as the desktop metaphor. A further advantage of the binaural system, often referred to as the "cocktail party effect" [10], [16], is that it improves the intelligibility of sources in noise and enhances the segregation of multiple sound sources. This effect could be critical in applications involving encoded information as in scientific "visualization," using the acoustic representation of multi-dimensional data [4], [28], or the development of alternative interfaces for the visually impaired [15], [28]. Another aspect of auditory spatial cues is that, in conjunction with other modalities, they can act as a potentiator of information in the display. That is, visual and auditory cues together can reinforce the information content of the display and provide a greater sense of presence or realism in a manner not readily achievable by either modality alone [1], [7], [11], [26], [29], [31]. This phenomenon will be particularly useful in telepresence applications, such as advanced teleconferencing environments, shared electronic workspaces, or monitoring telerobotic activities in remote or hazardous situations. Thus, the combination of direct spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use.

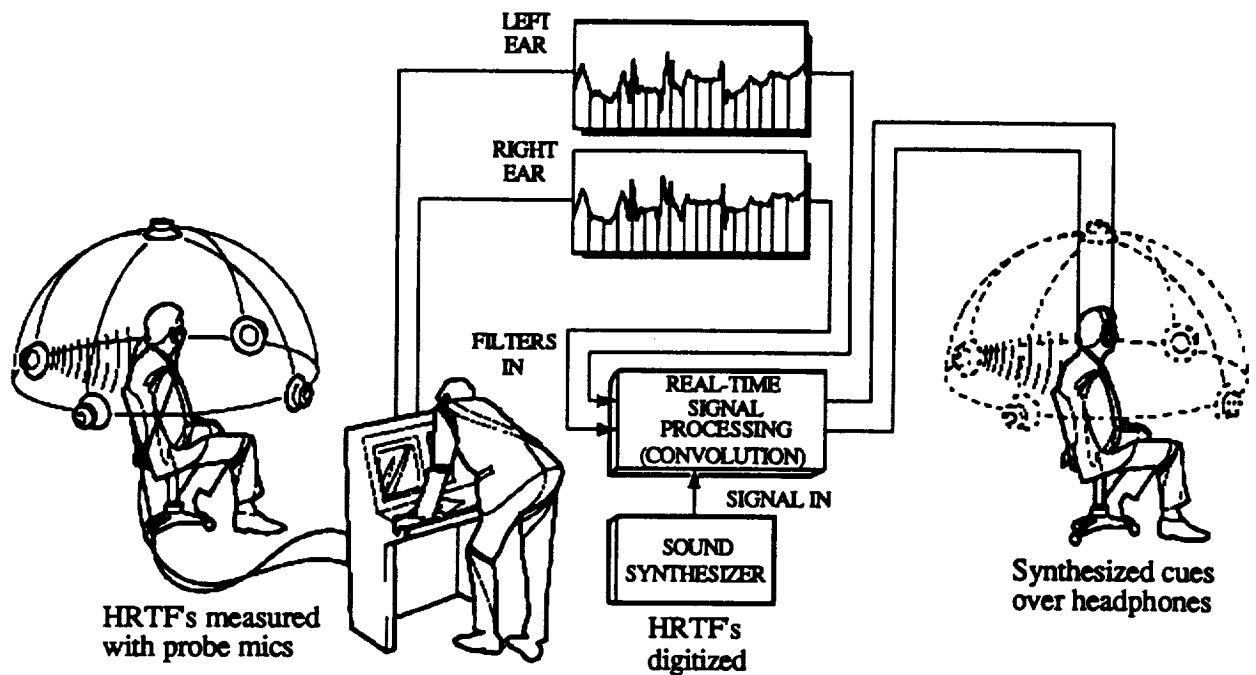


Figure 1: 3D Auditory Display; Synthesis Technique

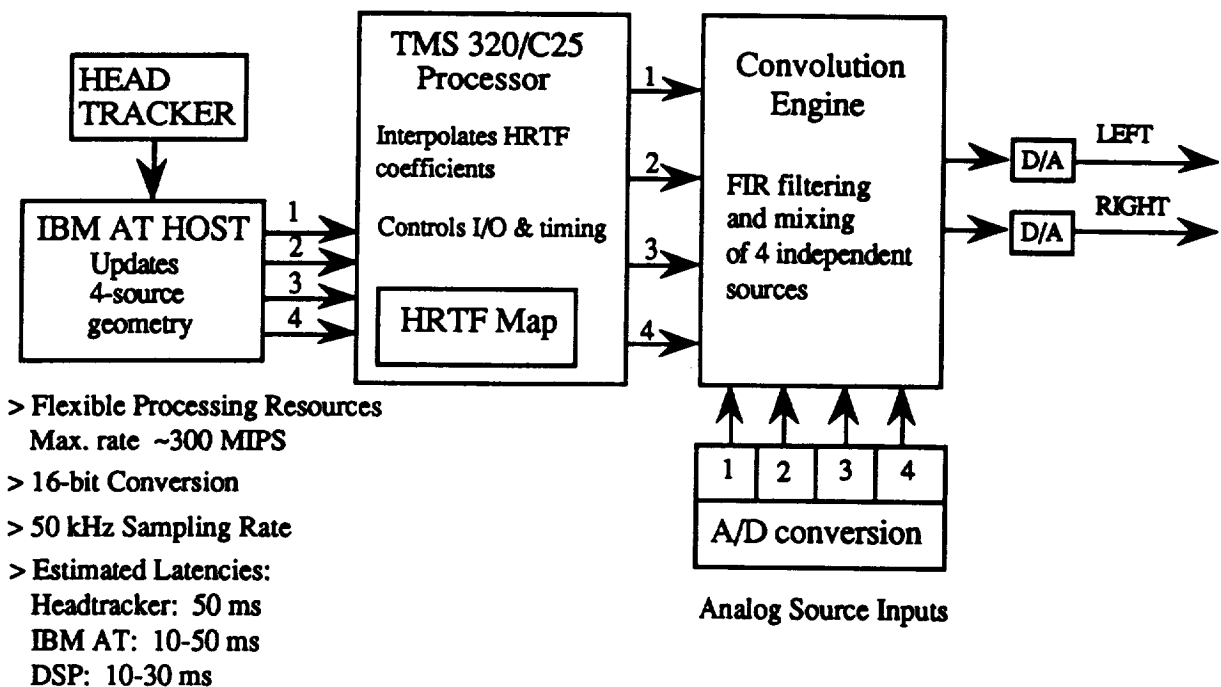


Fig. 2: The Convolvotron

### Implementing Three-Dimensional Sound



Localized acoustic cues could be realized with an array of real sound sources or loudspeakers [9], [13]. An alternative approach, recently developed at NASA-Ames Research Center, generates externalized, three-dimensional sound cues over headphones in real time using digital means [32], [33]. This type of presentation system is desirable because it allows complete control over the acoustic waveforms delivered to the two ears and the ability to interact dynamically with the virtual display. The synthesis technique, illustrated in Figure 1, involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the ear-canals of individual subjects (see [36], [3]). The advantage of this technique is that it preserves all of the interaural temporal and level differences over the entire spectrum of the stimulus, thus capturing the effects of filtering by the pinnae which are critical for the veridical simulation of externalized sound sources.

In the real time system, the Convolvotron, up to four moving or static sources can be simulated in a head-stable environment by digital filtering of arbitrary signals with the appropriate HRTFs. Motion trajectories and static locations at greater resolutions than the empirical data are simulated by linear interpolation of the four nearest measured transforms. Also, a simple distance cue is provided via real time scaling of amplitude. Figure 2 shows the functional components of the Convolvotron system designed by Scott Foster.

Such an interface not only requires the development of special-purpose display technology, it also necessitates the careful psychophysical evaluation of listeners' ability to accurately localize the virtual or synthetic sound sources. The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience will be duplicated. A recent study [36] confirmed the perceptual adequacy of the basic technique for static sources for experienced subjects localizing stimuli in the free-field compared with stimuli synthesized from their own HRTFs. Source

azimuth was synthesized nearly perfectly for all listeners while synthesis of source elevation was less well-defined, e.g., more variable with a compressed range of responses. Elevation was also the source of the most obvious individual differences in localization for both free-field and synthesized signals.

Unfortunately, measurement of each potential listener's HRTFs may not be possible in practice. It may also be the case that the user will not have the opportunity for extensive training. Thus, a critical research issue for virtual acoustic displays is the degree to which the general population of listeners can obtain adequate localization cues from stimuli based on non-individualized transforms. Preliminary data [34] from

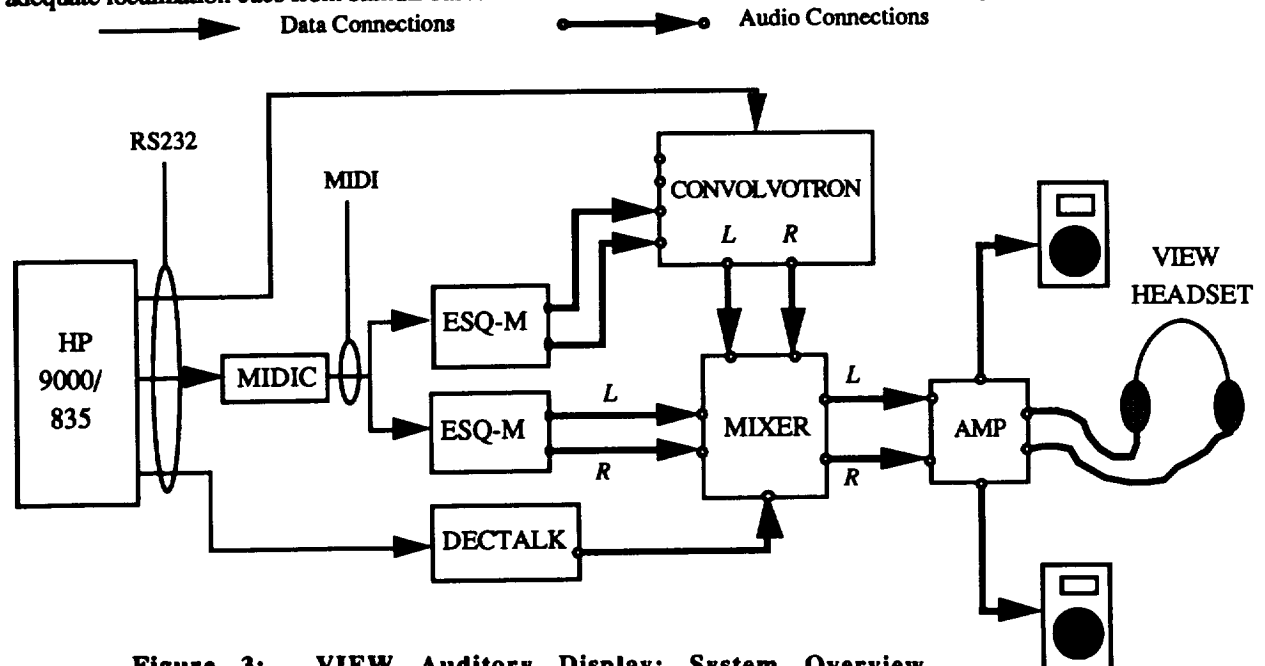


Figure 3: VIEW Auditory Display; System Overview

three experienced subjects suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance was only somewhat degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer. Further, the fact that individual differences in performance, particularly for elevation, can be traced to acoustical idiosyncrasies in the HRTF spectra, suggests that it may eventually be possible to create a set of "universal transforms" by parametric modeling techniques (e.g., [24]), principal components analysis, or perhaps even enhancing the spectra of empirically-derived transfer functions (e.g., [14]).

### VIEW Sound System Architecture

While perceptual studies of individual sensory modalities are clearly needed, it is also important to examine the role of sensory interaction. NASA-Ames' VIEW system provides the opportunity to implement localized auditory icons and assess their contribution to an integrated spatial display. Briefly, VIEW is a multisensory display environment which allows the user to explore and interact with a 360-degree synthesized, or remotely-sensed, world using a head-mounted, wide-angle stereoscopic display controlled by operator position, voice, and gesture. More detailed descriptions of the VIEW visual and gestural displays can be found in [17] - [19].

The VIEW auditory display subsystem allows audio cues responsive to both discrete events and continuous data changes to be designed and linked to arbitrary events and data flows in VIEW scenarios. Refer to the system overview diagram, Figure 3, for the following discussion.

Development of the initial binaural display capability based on MIDI (Musical Instrument Digital Interface) sound-synthesis technology began in 1987. More recently, true spatial cueing was added to the system with the integration of the Convolvotron. The auditory display subsystem, like most of the VIEW system, is currently implemented with a Hewlett Packard HP9000/835 computer. Two Ensoniq ESQ-M synthesizer modules handle the actual production of audio cues, supplemented with a Digital Equipment Corporation DECTalk speech synthesizer. MIDI protocol is used to communicate between the HP host and the ESQ synthesizers. A Hinton Instruments MIDIC interface converts 19.2 Kbps RS232 signals from the HP into 31.25 Kbps 5 mA. current-loop signals that are required by the MIDI standard.

Each ESQ synthesizer has two outputs. One ESQ's output pair is mapped directly into the VIEW system's left and right audio channels. Up to eight independent (polytimbral) voices, mixed to the stereo output, may be played through this synthesizer. The second ESQ's output pair is patched into the Convolvotron. As described above, this device is capable of synthesizing, in real time, an apparent three-dimensional location for up to four independent audio inputs. In this case, since only two channels of sound are available from the ESQ, only two of the Convolvotron's inputs are used. The simple stereo pair from the first ESQ, and the 3D-imaged output from the Convolvotron, are mixed, amplified, and sent to headphones integrated into the VIEW headset, or optionally, to room speakers.

The central software component of the auditory display is the cue driver. As in all VIEW applications, the software is written in C in a Unix environment. Without delving too deeply into its implementation details, it can be described as consisting of an event scheduler, a MIDI, speech, and Convolvotron event generator, and a VIEW/Auditory Display rendezvous mechanism. It also handles several housekeeping chores, such as loading cue files, initializing the MIDI interface, the synthesizers and the Convolvotron, downloading patch files to the synthesizers, and making sure all is quiet before a VIEW scenario exits.

Up to ten (monophonic) auditory "objects" may be displayed simultaneously. If a more complex, or polyphonic, sound is desired, several voices can be assigned to the same icon and the number of possible simultaneous objects is reduced accordingly. In general, the basic sound signature or identity of an individual object or "icon" derives from the particular ESQ patch assigned to it. Since this technology was developed for music synthesis, one can often think of a patch as having the attributes of a particular musical instrument. However, some "environmental" sounds analogous to sound effects, e.g., footsteps or explosions, are also possible.

Custom ESQ patches may be designed off-line using the front-panel capabilities of the synthesizer, a patch editor/librarian software package, or by selecting from collections of commercially-available pre-

programmed patches. Icons may take advantage of one or more of the controllable parameters made available by the ESQ. These include oscillator frequency, filter cutoff frequency, amplitude level, and stereo pan position. Any of these can be modulated in real time and associated with events or information flow in a VIEW display scenario. One of the advantages of the ESQ, and the main reason for its choice during the specification of this system, is that it allows access to these parameters through standard MIDI controllers; many synthesizers require the use of "system-exclusive" messages to achieve this level of control. Designing the cue driver around standard MIDI controllers makes it less system-dependent; as more synthesizers adopt this level of control (and this seems to be the trend), the auditory display may be readily adapted to them.

Because of the limited outputs of the ESQ (two per synthesizer), only two of the ten icons may be assigned specific locations via the Convolvotron at any one time. Alternatively, using all four ESQ outputs, up to four simultaneous icons could be independently localized. The current configuration was a compromise solution which traded localized cues for an increase in the number of possible icons. A future solution would be to adapt the system to a synthesizer which has independent outputs for each voice. Also, integrating a digital-sampling device would be useful for presenting the kinds of sounds that Gaver [23] advocates in his notion of "everyday listening." At the time we began developing the system, digital samplers tended to be expensive and allowed very little real time control over the acoustic parameters of the sounds. Since a major goal of the display was to allow continuous control over the icons' acoustic structure, we opted for a more standard and inexpensive synthesizer with a relatively well-developed MIDI implementation. [See [8] for a useful discussion of the pros and cons of various MIDI devices.]

### Editing & Display Capabilities

Cues or icons are designed and refined with ACE, the Auditory Cue Editor. ACE is a stand-alone program, which makes it unnecessary to activate the entire VIEW system merely to work on auditory cues. Multi-level menus, interactive prompting, and extensive syntax checking aid the user in designing complex auditory cues with relative ease.

ACE is composed of four basic sections, organized as independent screens, each with its own menu of commands.

On the Main Screen, cues may be created, deleted, loaded, saved, and named (an important function; all rendezvous between VIEW events and data are made through the names of cues as specified on the Main Screen). The Main Screen may also be used to specify certain basic parameters, e.g., synthesizer patch number and localization method (convolved or simple stereo). In addition, a "play" command allows quick, interactive audition of cues during their design.

The Sound Event Editing Screen allows the construction of the main body of a cue, which is in the form of a list of time-ordered MIDI, speech, or Convolvotron localization commands. While entire pieces of music could theoretically be entered here, note by painstaking note, this is usually not the case. Single notes, chords, or short sequences of notes and chords are the most common items entered on this screen. This is because very simple events (with carefully chosen and distinct timbres) are often all that is needed for basic cueing functions. Even the more complex auditory icons generally have fairly simple event lists, since most higher-level display capabilities result from linkage to continuous data streams and response to real-time changes of those data streams.

The Modulation Editing Screen enables the connection and scaling of incoming data streams to several different MIDI modulators as well as Convolvotron sound-source position coordinates. This makes it possible to have an arbitrary data value produced by the VIEW system displayed as a proportional deflection in a variety of auditory parameters, such as pitch, timbre, or apparent three-dimensional location. The data structure has provisions for incorporating nonlinear mappings between incoming data and an auditory parameter, a feature which can be very important in evaluating the perceptual consequences of a particular icon. Specifications of modulation behavior composed on this screen are given textual names, which enables rendezvous with the appropriate VIEW-generated data streams.

Patches, as mentioned above, are a very important part of an auditory icon; they define the basic sound that the synthesizer produces, and the manner in which it will react to incoming controls sent by the cue driver.

The Patch Maintenance Screen allows the uploading and downloading of individual patches and complete patch banks between the host system and the synthesizers. In this way, a particular set of sound programs can be directly associated with a set of auditory cue definitions.

### **Application in the Virtual Environment: Telerobotic Control**

The VIEW telerobotic scenario was designed to illustrate the capability of telepresence, i.e., the manipulation of objects or interaction with persons or objects remote from one's location, that a virtual environment makes possible. In this scenario, the visual and kinematic characteristics of a Puma robotic arm are modeled with a high degree of accuracy. The scenario participant may, upon donning the VIEW stereoscopic head-mounted display, align his or her arm with that of the lifesize model. The participant's arm and the modeled robotic arm may then be "coupled," which simply means that the robotic arm will move, to the extent of its kinematic capabilities, in correspondence to the movement of the participant's arm. During the coupled mode, an end-effector with a vise-like gripping apparatus may be opened and closed merely by opening and closing the hand.

This graphic computational model of a robotic arm is meant to test the efficacy of the intuitive mapping of control between machine and human counterpart. The success of this mapping is tested by assigning a simple task to be completed by the telerobotic participant. In the foreground of the scenario, a "circuit card" is plugged into a slot on a "task board." The participant is instructed to remove the circuit card and replace it with another one which is just off to the side of the task board. This entails coupling with the robotic arm, maneuvering the end-effector into position so that the two jaws surround the edge of the circuit card, closing the end-effector jaws around the card to grasp it, and pulling it out and away from the task board. Once this is completed, the replacement circuit card must be grasped in a similar manner, lined up exactly, and inserted into the slot.

With perfect telepresence, this task could be accomplished with little more difficulty than if one were using one's own hand and a real task board and circuit card. However, factors such as slower-than-ideal graphic refresh rates, lower-than-ideal contrast and focus in the VIEW display, etc., conspire to make the precision manipulation required somewhat difficult. In a situation such as this, auditory feedback can make an important difference, particularly with the current paucity of good haptic or force feedback display systems.

At the simplest level, auditory feedback is used to indicate the occurrence of discrete events in the scenario. For example, many commands and actions in VIEW are initiated by hand gesture. A VPL "Data Glove" reports finger positions to the host computer, which examines those positions for correspondence to any of several pre-defined gestures, such as "single-finger point," or "fist." When one of these gestures is detected by the host, a sound is made by the auditory display to indicate gesture recognition.

Other simple auditory cues fall into the category of "reality-mimicry," or sound effects. In the telerobotic scenario, bumping an end-effector into a "solid" object in the virtual (or teleoperated) world causes a "bump" sound to be produced. Since direct force feedback is not yet available in the VIEW system, this form of audio display is particularly critical, as it warns of a situation which could cause damage to a real-world robotic arm or to objects with which it is colliding. At a more mundane level, this sort of sound effect enhances the sense of presence; objects tend to make a sound when they collide in the real world, so it is reasonable to expect them to do so in a virtual environment.

Audio feedback that supplements or replaces force feedback is not limited to mimicry of collision sounds or similar sound effects. Force can be represented as a continuum by changing one or more sound parameters in correspondence to the force's intensity. This type of display is utilized for a special circumstance in the telerobotic scenario. If the scenario participant attempts to force the replacement circuit card into the task board without orienting it correctly, a force-reflection display called "push-through" is initiated. It starts out as a soft, steady tone that gets louder, brighter (higher harmonics are let through the filter), and more frequency modulated the harder the participant pushes on the misaligned card. In this way, a potentially damaging increase in user input is signalled by an increasingly harsh and strident auditory warning.

Taking this idea one step further, not only force, but any arbitrary continua of data may be displayed. Perhaps the most successful use of auditory feedback in the telerobotic scenario comes into play while the participant attempts to guide the replacement circuit board into the target socket. As the board reaches a certain proximity to the socket, a cue initiates consisting of two sustained tones; the pitch of one of the tones is deflected with respect to the other by an amount proportional to the distance between the circuit card and its slot. As the card nears the slot, the two pitches come closer together (which is readily perceived due to the obvious decrease in the beat frequency produced by the increasingly adjacent pitches); at distance zero, the tones are in unison. The cue functions as an auditory "rangefinder," and greatly facilitates the proper positioning of the card in its slot.

Card orientation, which is also crucial to the completion of the replacement task, could be represented by some other continuous audio parameter, such as depth of frequency modulation. With careful selection and scaling of the modulated sound parameters, two continua (e.g., proximity and orientation) could be monitored simultaneously at a very intuitive level. Our work to date has not explored multiple-simultaneous displays of continuous data, but it is an intriguing area for further research.

### **Auditory Design Principles**

The telerobotic scenario has served as an excellent test of the capabilities of the auditory display. While formal experimentation has yet to be done, it has also provided a rich environment for the discovery of certain basic guidelines for the design of auditory icons and the development of an auditory symbology.

Practical experience has shown that the most effective cues are simple cues. Long sequences or elaborate clusters of tones not only tend to clutter the auditory display, they can increase the load on cognitive processing and memory required to interpret the information, and in the long run, become downright annoying. Imagine a telephone that played "Three Blind Mice" every time a call came in. It would be only a short time before the exactly-repeating melody became maddening. The simple bell or digital chirp of a telephone manages to get attention without engaging the "music critic" part of one's cognition. Similarly, a "thud" sound suffices to signal a bump in a virtual world; it is not necessary to have a speech synthesizer say "You have bumped into something" at each and every collision. While these may be extreme examples, the basic principle holds; an auditory icon should be as simple as possible.

The need for simplicity is even more critical when several cues occur in close proximity to each other. For instance, in the telerobotic scenario, a gesture-recognition cue might be followed immediately by a sound that indicates movement of the jaws of the end-effector. If the jaws were then to close over the circuit board, a "board-grasped" cue would result. These three cues can occur in rapid succession, so they must be of short duration for the correct sequence of events to be properly represented.

This situation also points out the need for carefully choosing the sound signatures or patches which form the fundamental units of an auditory symbology. Patch design, including spectral content, amplitude and filter envelopes and various special effects, is the chief distinguishing feature of a simple icon. The best way to make an icon recognizable is to give it a distinctive sound. Much effort in the design of auditory icons is therefore concentrated in selecting or building an appropriate synthesizer patch.

As noted before, guidelines can be derived from the fields of psychoacoustics, music, and perceptual psychology. As illustrated in the proximity cue in the telerobot scenario, the close tuning of two pitches is a continuous parameter to which the human ear is very sensitive. However, the amplitude modulation or beat frequency which signals the change in proximity will only occur for a limited range of frequencies which must be considered when mapping the distance data to the difference in pitch. In developing a symbology, one can also take advantage of what one might think of as "natural" or metaphorical mappings even if a literal sound, such as a "bump," is not possible. For example, the "push-through" cue described above clearly signals an increasing violation of the allowable forward movement when inserting the task board at an incorrect orientation by a harsh sound which increasingly "violates" the ears. To minimize cognitive effort, it is important to build meaning by the relationships between icons as well. In the telerobot scenario, icons which provide feedback for related gestures have similar timbres that are distinguished by their temporal structure. For example, larger changes in pitch, at the level of short sequences, are used (much like the familiar two-chime doorbell). This particular type of icon has the virtue of being reversible like a short musical motive; a "grasp" gesture is represented by a high note followed by a slightly lower note. The complementary "release" gesture is the same two notes, only in reverse order.

As much as possible, this relationship between sound and meaning should remain consistent throughout a display system. Thus, in VIEW, the cues which provide feedback for the various gestures remain the same across the different types of display scenarios that have been developed.

Careful consideration of the possible interactions between icons will be particularly important when auditory cues must be presented simultaneously as in the combination of orientation and proximity cueing described above. Principles of acoustic perceptual organization, such as the Gestalt principles elaborated by Bregman [6], will provide important guidelines. For example, different acoustic objects may be defined by different auditory streams. Streaming is determined by such features as frequency separation, timbre, rate or tempo, spatial location, and "common fate" or the tendency of spectral components to be grouped according to similar frequency or temporal patterns.

### **Other VIEW Scenarios**

Other examples of display scenarios which have been implemented in the VIEW system include the Extra-Vehicular Activity (EVA) Visor and a Computational Fluid Dynamics (CFD) data visualization. The EVA Visor is a concept for a helmet-mounted, three-dimensional dataspace which can be accessed by an astronaut during repair or inspection activities while outside the space station. In the scenario, several types of display windows can be used, including life-support system status, a "cuff checklist" of tasks to be completed, repair schematics, and a three-dimensional "map cube," which represents the entire EVA scenario as a miniature, manipulable cube, with which the astronaut can establish his or her position and orientation with respect to the other vehicles or objects in the scenario. Currently, auditory cues for the EVA Visor include a set of gesture recognition cues identical to those in the telerobot scenario (in keeping with the principle of transference of a learned auditory vocabulary between virtual worlds, when possible). Warning cues signal situations which might endanger the astronaut's safety, such as impending depletion of life support resources. A special sound effect cue indicates when MMU (the rocket-backpack vehicle which enables the astronaut to maneuver during EVA) thrusters are firing. Finally, activation of the various windows listed above is heralded by corresponding audio signals. Another cue, not yet implemented, which would be useful in the EVA display, is an orientation beacon which allows the astronaut to continuously monitor the location of the space station by means of a localized auditory icon to minimize disorientation in the absence of visual and gravitational referents.

The CFD data display visualizes the fuel-flow around the LOX (liquid oxygen) post of the main shuttle engine. Features include the ability to "fly through" the data, viewing it from different viewpoints including inside the fluid flow, and "grabbing" and scaling the data up or down to examine its finer or coarser features. Although not yet implemented, a potentially useful auditory visualization cue might be to "attach" auditory icons to one or more particles in the flow, and thus follow their progress as they interact with the structures of the shuttle engine.

In developing the VIEW auditory display, we have attempted to provide a flexible and general-purpose system which takes advantage of our knowledge of perceptual abilities as much as possible. The hardware architecture and software is designed to be applicable to a wide variety of display configurations and to allow a consistent approach to the design of auditory symbologies based on knowledge gleaned from music, psychoacoustics, and perceptual psychology.

### **References**

- [1] Begault, D.R. & Wenzel, E.M. (1990) Techniques and applications for binaural sound manipulation in man-machine interfaces. NASA Technical Memorandum No. TM102279, In Press.
- [2] Blattner, M.M., Sumikawa, D.A., & Greenberg, R.M. (1989) Earcons and icons: Their structure and common design principles. *Hum.-Comp. Interact.*, 4, 11-44.
- [3] Blauert, J. (1983) *Spatial Hearing*. The MIT Press: Cambridge, MA.
- [4] Bly, S. (1982) *Sound and computer information presentation*. Unpublished doctoral thesis (UCRL-53282) Lawrence Livermore National Laboratory and University of California, Davis, CA.

- [5] Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys., Washington, D.C., 1-11.
- [6] Bregman, A. (1981) Asking the "what for" question in auditory perception. In Kubovy & Pomerantz (Eds.), *Perceptual Organization*, Lawrence Erlbaum Associates: Hillsdale, NJ.
- [7] Brown, M., Newsome, S., & Glinert, E. (1989) An experiment into the use of auditory cues to reduce visual workload. Proceedings of CHI'89, ACM Conference on Human Factors in Computing Systems, 339-346.
- [8] Buxton, W., Gaver, W., & Bly, S. (1989) The use of non-speech audio at the interface. Tutorial #10, CHI'89, ACM Press: New York.
- [9] Calhoun, G.L., Valencia, G., & Furness, T.A. III (1987) Three-dimensional auditory cue simulation for crew station design/evaluation. Proc. Hum. Fac. Soc., 31, 1398-1402.
- [10] Cherry, E.C. (1953) Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am., 22, 61-62.
- [11] Colquhoun, W.P. (1975) Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. Hum. Fac., 17, 425-437.
- [12] Deatherage, B.H. (1972) Auditory and other sensory forms of information presentation. In H.P. Van Cott & R.G. Kincade (Eds.), *Human Engineering Guide to Equipment Design*, (rev. ed.), Washington, DC: U.S. Government Printing Office, 123-160.
- [13] Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.
- [14] Durlach, N.I. & Pang, X.D. (1986) Interaural magnification. J. Acoust. Soc. Am., 80, 1849-1850.
- [15] Edwards, A.D.N. (1989) Soundtrack: An auditory interface for blind users. Hum. Comp. Interact., 4, 45-66.
- [16] Egan, J.P., Carterette, E.C., & Thwing, E.J. (1954) Some factors affecting multichannel listening. J. Acoust. Soc. Am., 26, 774-782.
- [17] Fisher, S.S. (1986) Telepresence master glove controller for dexterous robotic end-effectors. *Advances in Intelligent Robotics Systems*, D.P. Casasent (Ed.), Proc. SPIE, 726.
- [18] Fisher, S.S., McGreevy, M.W., Humphries, J., & Robinett, W. (1986) Virtual environment display system. ACM Workshop on Interactive 3D Graphics, Chapel Hill, NC.
- [19] Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. Proc. Hum. Fac. Soc., 32, 91-95.
- [20] Foley, J.D. (1987) Interfaces for advanced computing. Sci. Amer., 257, 126-135.
- [21] Furness, T.A. (1986) The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., 1986 (1), 48-52.
- [22] Garner, W.R. (1949) Auditory signals. In *A Survey Report on Human Factors in Undersea Warfare*, Washington, D.C.: National Research Council, 201-217.
- [23] Gaver, W. (1986) Auditory icons: Using sound in computer interfaces. Hum.-Comp. Interact., 2, 167-177.

- [24] Genuit, K. (1986) A description of the human outer ear transfer function by elements of communication theory. Proc. 12th ICA (Toronto), Paper B6-8.
- [25] Loomis, J.M., Hebert, C., & Cicinelli, J.G. (1990) Active localization of virtual sound sources. Submitted to J. Acoust. Soc. Am.
- [26] O'Leary, A. & Rhodes, G. (1984) Cross-modal effects on visual and auditory object perception. *Perc. & Psychophys.*, 35, 565-569.
- [27] Patterson, R.R. (1982) Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority Paper No. 82017, London.
- [28] Smith, S., Bergeron, R.D., & Grinstein, G.G. (1990) Stereophonic and surface sound generation for exploratory data analysis. Proceedings of CHI'90, ACM Conference on Human Factors in Computing Systems, 125-132.
- [29] Sorkin, R.D., Wightman, F.L., Kistler, D.J., & Elvers, G.C. (1989) An exploratory study of the use of movement-correlated cues in an auditory heads-up display. *Hum. Fact.*, 31, 161-166.
- [30] Sutherland, I.E. (1968) Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., 33, 757-764.
- [31] Warren, D.H., Welch, R., & McCarthy, T.J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perc. & Psychophys.*, 30, 557-564.
- [32] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988a) Development of a three-dimensional auditory display system. *SIGCHI Bulletin*, 20, 52-57.
- [33] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988b) A virtual display system for conveying three-dimensional acoustic information. *Proc. Hum. Fac. Soc.*, 32, 86-90.
- [34] Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. (1988c) Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Amer.*, 84, S79.
- [35] Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Amer.*, 85, 858-867.
- [36] Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. *J. Acoust. Soc. Amer.*, 85, 868-878



# FAST<sup>1</sup>: A Multi-Processed Environment for Visualization of Computational Fluid Dynamics

Gordon V. Bancroft  
 Fergus J. Merritt  
 Todd C. Plessel  
 Paul G. Kelaita  
 R. Kevin McCabe  
 Al Globus

Sterling Federal Systems Inc.  
 1121 San Antonio Road  
 Palo Alto, California 94303

## Abstract

Three-dimensional, unsteady, multi-zoned fluid dynamics simulations over full scale aircraft is typical of problems being computed at NASA Ames' Numerical Aerodynamic Simulation (NAS) facility on CRAY2 and CRAY-YMP supercomputers. With multiple processor workstations available in the 10-30 Mflop range, we feel that these new developments in scientific computing warrant a new approach to the design and implementation of analysis tools. These larger, more complex problems create a need for new visualization techniques not possible with the existing software or systems available as of this writing. These visualization techniques will change as the supercomputing environment, and hence the scientific methods employed,<sup>1</sup> evolve even further.

Visualization of computational aerodynamics requires *flexible*, *extensible*, and *adaptable* software tools for performing analysis tasks. *Flexible* means the ability to handle a diverse range of problems. *Extensible* means the ability to interact at all levels of the software hierarchy, either through existing built-in functionality or through the implementation of custom "plug-in" modules. *Adaptable* means the ability to adapt to new software and hardware configurations through the use of modular structured programming methods, a graphics library standard, and the use of common network communication protocols (like UNIX sockets) for the distribution of processing.

This paper discusses FAST (Flow Analysis Software Toolkit), an implementation of a software system for fluid mechanics analysis that is based on this approach.

## Background

Computational Fluid Dynamics (CFD), involves the use of high speed computers to simulate the characteristics of flow physics. Computational aerodynamicists use CFD methods and solvers to study subsonic, supersonic, transonic and hypersonic (compressible) regimes of flight, in addition to studying incompressible problems within particular systems. Examples of ongoing studies on full-scale aircraft configurations at NASA Ames include the Space Shuttle, F16, and the Aerospace Plane. Specialized areas of research include jet-engine turbine flow, VSTOL (Vertical / Short Take Off and Landing) and ground effect research, and even flow through an artificial heart. Basic CFD research involves unsteady flow phenomena like vortex shedding and turbulence modelling.

A flow solver running on a supercomputer must handle input files (finite difference grids, ref. 7,15,16) that are typically very large. For example, the number of xyz triplets (each represented by three eight-byte floating point numbers) in a 100 x 100 x 100 grid yields a 24 Mbyte file. If complexity is added, or the grid resolution (density of points) must be raised for flow solving to yield acceptable results, the files grow

---

<sup>1</sup> FAST (Flow Analysis Software Toolkit) Developed by Sterling Federal Systems Inc. under contract to NASA Ames Research Center NASA Contract #NAS2-11555.

proportionally in size. Once the solver has been run, there are from five to eight variables for each grid node, again, each represented by an eight-byte floating point number. For the 24 Mbyte example, five variables for each grid point yields a 40 Mbyte raw data file. This is a total of 64 Mbytes (grid plus the solution) for this example. The F16 mentioned previously, which consists of 29 grid zones, is over 108 Mbytes worth of data! *(Note: On the workstations these become four-byte IEEE format floating point numbers making the files about half this size)*

Two examples of grid generation programs are:

<b>3DGRAPE</b>	3-dimensional grids about anything by Poissons Equation (Sorensen)
<b>IZ</b>	Interactive zoner (Cordova)

A list of commonly used flow solvers are:

<b>ARC2D</b>	Ames Research Center 2-dimensional solver (Pulliam)
<b>ARC3D</b>	Ames Research Center 3-dimensional solver (Pulliam)
<b>TNS</b>	Transonic Navier Stokes solver (Flores)
<b>CNS</b>	Compressible Navier Stokes solver (Flores)
<b>PNS</b>	Parabolized Navier Stokes solver (Chausee)
<b>INS3D</b>	Incompressible Navier Stokes solver (Kwak)
<b>TWING</b>	Transonic Wing solver (Thomas)

Programs available for visualization of CFD data sets are:

<b>PLOT3D</b>	A command line driven Fortran program that computes CFD quantities (Buning [7])
<b>SURF</b>	Allows for the rendering of smooth, wireframe, and function mapped surfaces with a more interactive interface (Plessel[8])
<b>GAS</b>	Combines graphics generated from PLOT3D and SURF and allows animations to be created and recorded (Merritt[9])
<b>RIP</b>	A program for interactive particle tracing (Rogers[19])

### **FAST Overview**

The software cycle for the creation and analysis of computational fluids results could be reduced to the following conceptual model:

- Data generation (*Flow solving*)
- Data manipulation (*The original data may need to be filtered or transferred*)
- Data abstraction (*A graphical object is defined using the data*)
- Data rendering (*Viewing on a workstation*)
- Data interpretation (*analysis*)
- Feedback (*Perhaps go back to previous phases*)

A problem with the existing CFD software is that it takes a non integrated approach to dealing with the different steps of the CFD process. The grid generation and flow solver programs are involved in the data generation phase. The visualization software is part of the abstraction, rendering and analysis phases. The various programs present the user with different interfaces, and there is little attention paid to the data manipulation and feedback steps. In the current system, large data sets flow from one step to another from disk to ram and back to disk (perhaps from one computer to another), taking on different file formats along the way.

The design criteria for FAST were:

- Minimize the data path in the CFD process
- Provide a consistent user interface
- Allow for quick user feedback
- Provide an extensible software architecture
- Provide a quick path through the CFD process
- Provide libraries and tools so that application modules could be added easily
- Isolate 3D viewing tasks from the application module programmer

In order to achieve these design goals FAST has evolved into collection of programs that communicate via Unix sockets with a central hub process that manages a pool of shared memory. A fundamental data type is loaded or generated and stored into shared memory (data generation and manipulation), a collection of programs (modules) operate on data and produce additional data (objects) that are also placed into shared memory (data abstraction). The objects are rendered using the fast viewing system (data rendering). Data is analyzed by additional modules or visual inspection (data analysis). Depending on the results of the analysis the user changes input to any of the previous modules (feedback). In addition there is a collection of libraries and utilities that are used to build the application modules.

The use of shared memory reduces the flow of data in the system. The use of a viewing process relieves the burden of three dimensional interactive viewing from the application programmer. The fact that the fundamental data type(s) reside in shared memory makes it easy to make changes based on the feedback obtained from the analysis phase. Finally the use of FAST libraries and utilities makes it easy to add new modules.

We are aware of other scientific visualization packages and visualization capabilities in existence and/or under development. These include visual programming examples like CONMAN (Silicon Graphics[3]) and AVS (Application Visualization System, Stardent Computer[4]), and other scientific visualization environments like MPGS (Multi-Purpose Graphics System, Cray Research), and the Personal Visualizer (Wavefront), as well as 'scripting' languages like PVWAVE (Precision Visuals), IVIEW (Intelligent Light), and VISAGE (Visual Edge) to name a few. While FAST is built specifically around the research tasks involved in CFD analysis, these other environments and packages typically take a much more generalized approach towards visualization, for the obvious reason that CFD research is a relatively small part of their intended audience. These systems and environments often require a certain level (a 'power' user, visual programmer, or animation /rendering expert) of skill with computer graphics above and beyond the level of the typical CFD scientist. In researching these other more general approaches, we have discovered that the results (data) get handed off at some point to the 'power' user (or perhaps even computer graphics group or expert) and this person (or group) creates the animations, films or videos. FAST is built around a model where the scientist is the first and last person in the data chain and FAST is a toolset for his environment. This is not meant as a criticism of these other approaches, as the need for generalization dictates the need for this other level of user. It is our belief, though, that the techniques used in FAST presented in this paper would also apply and be very useful in the more general environments.

Graphics, CPU, and memory handling performance were key considerations in the FAST design and development process. For graphics, a base-line level of what is commonly termed (but undefined) as "real-time" had to be established and agreed upon as acceptable. This was determined to be a minimum of 3 frames/sec for a typical 10-20 Mbyte problem (techniques used for rendering would determine the problem size in this range). This base line frame rate was determined to be essential in visualization of fluid mechanics for understanding the dynamics of the simulations. For the development platform, the Silicon Graphics 4D220/GTX (16 Mbytes memory) this goal was reached and we are very pleased with the current performance level. The targeted platform, the Silicon Graphics 4D320/VGX, is expected to have even higher levels of cpu and graphics performance[18].

We have implemented in FAST new techniques and capabilities non-existent in the previous tools and expanded on others. For example, the colormap editing capabilities were enhanced to include banded, spectrum, dynamic, contour, striped, and two-tone function mapping. Surface rendering includes the ability to 'sweep' planes through the data either grid oriented, arbitrarily oriented, or a contour surface (isosurface).

Enhanced titling and labelling features include the use of postscript type fonts and symbols, where typeface, font point size, and style can be specified. The animation capability is substantially enhanced beyond what was available in GAS (Graphics Animation System[9]). These enhancements include greater control by allowing the ability to edit scenes, views, and objects. Another capability allows for separate scenes to be rendered in separate windows giving the scientist/user even more flexibility and animation control.

### **FAST Architecture**

Each separate process communicates through the FAST Hub while managing shared memory and communicating using standard Berkeley UNIX Interprocess Communication (IPC[11]).

#### **Hub**

The central process of the FAST environment is the Hub module. The Hub module invokes and shuts down the FAST modules yet its main function is to process requests sent by the modules. These requests might be to allocate a segment of shared memory and return the shared memory id, or to delete a shared memory segment. Since the Hub process is always running as long as FAST is active, the data allocated through the Hub remains accessible even when the original process which requested it is terminated. The Hub module is essentially transparent to the user, in that it has no panels.

#### **FAST Central**

This is the central module for processing, from the users perspective. This is where the graphical data pool generated by other modules is managed and interactively viewed. FAST Central, unlike other FAST modules, runs continuously while FAST is up and running. Other modules can be spawned or shut down as they are needed from the FAST Central module. In addition FAST Central allows object attributes to be set (e.g. transparency, mirroring, line width), scene attributes to be set (e.g. lighting, color map editing, background color), viewing preferences to be set (e.g. toggle axis, mouse axis modes) as well containing the animation control panels. Animator is used to create and record smooth (spline interpolated) keyframe animation sequences.

#### **File Input**

The file input module loads pre-computed *PLOT3D* type grid, solution, and function files as well as ARCGraph[20] files into FAST's shared memory. It consists of three control panels. The file input panel is used to list file names and information and to load data into shared memory. The data sub-panel displays pertinent information about the previously loaded grids and solutions. The ARCGraph panel is used for handling this type of file input.

#### **CFD Calculator**

The CFD Calculator module allows the scientist to attach to the grid and solution data that has been loaded and to calculate a variety of scalar and vector functions for analyzing the computed solution. The Calculator has the appearance and functionality of a real programmable calculator but instead of operating on numbers it operates on *fields* of numbers (scalars) and *fields* of vectors.

Its basic operations (e.g., +, -, MAG, CURL), are applied to entire fields - either component-wise or vector-wise. For example, + applied to two scalar fields will produce a new scalar field of values that are the sums of the corresponding values of the two operand scalar fields. And LOG applied to a vector field will generate a new vector field by taking the logarithm of each component of the corresponding operand vector. In addition to component, scalar and vector binary operators there are also special operations such as GRADIENT, DIVERGENCE, DOT, and CROSS that apply to entire fields and produce new scalar or vector fields.

The scientist can select a range of active solution zones on which to operate and use the CFD Calculator to compute about 100 different built-in CFD scalar and vector functions such as Pressure, Enthalpy, Normalized Helicity, Velocity, and Vorticity [16]. These fields are stored in one of the Calculator's scalar or vector *registers*. The Calculator can then be *programmed* with formulas that operate on these fields and

produce new ones using the basic operations already mentioned. The CFD Data Panel is used to copy, move, delete, and display information fields (such as min-max) stored in the Calculator's registers. These features, and others, help make the CFD Calculator an interactive, powerful tool that the CFD scientist can use to compute important quantities for analyzing computed solutions.

### SURFER

The SURFace Extractor and Renderer module attaches to grids (loaded by the File Input Module) and scalar and vector fields (generated by the CFD Calculator) and renders grid surfaces as points, lines, vectors, or polygons. These *grid surface objects* are also stored in shared memory so they can be rendered in the FAST environment. The grid surfaces can show the grid geometry, for example, a lighted, Gouraud [2] shaded polygon surface of the Space Shuttle, or they can display the scalar data as function colored lines or polygons, or vector data as line vectors, vector heads, or polygon vector deformation surfaces (vector heads connected in a surface). Grid surface objects can represent grid geometries, scalar fields, and vector fields.

In addition to changing data types, surface rendering and other attributes, SURFER can sweep through all surfaces in a given grid direction. This creates a dynamic image showing even more features of the flow field.

### Titler

The Titler module is used to create high quality Postscript text suitable as titles for images in videos, slides, and movies. Title attributes include font, point size, position, color, drop shadows, and a snap-to-grid feature to make alignment easier. Like other graphical objects, *title objects* are stored in shared memory so they can be added to other scenes. Postscript fonts from other sources may be imported and created titles may be saved for later use.

### Isolev

Isolev performs three functions using a single algorithm. One, it draws surfaces of constant value in 3D scalar fields, i.e. isosurfaces. Two, it draws cutting planes function mapped by the scalar field of interest. Cutting planes may be at any angle, and are consistently oriented throughout a multi-zoned grid. Three, it draws vector field deformation surfaces originating at cutting planes or isosurfaces. Iso and deformation surfaces are lighted and smooth shaded. Both isosurfaces and cutting planes may be rendered as dots for improved performance. Interactive grid coarsening is available to improve interactivity. The user may also set up sweeps, where isolev automatically sweeps the isovalue (or cutting plane location) through all possible values, or within a user specified range. This can be used to get a feel for the entire volume. The marching cubes algorithm [Kerlick,13] is used to generate polygons. Level scalar fields are created to generate cutting planes function mapped by the scalar field of interest. Edge crossings, a faster algorithm, is used to generate points. A user selected vector field may be used to draw vectors originating at the crossing points.

### Tracer

The tracer module is used to compute particle traces and render them as vectors through the flow field. Tracer attaches to a grid and solution and allows the user to interactively select the point of release or rake[7] from which the traces are computed. The traces can either be computed forward or backward in time as well as allowing the user to selectively save traces. Once traces are saved, a delta time factor may be interactively adjusted through the panel to allow particle trace "cycling".

## Interactive Visualization Control

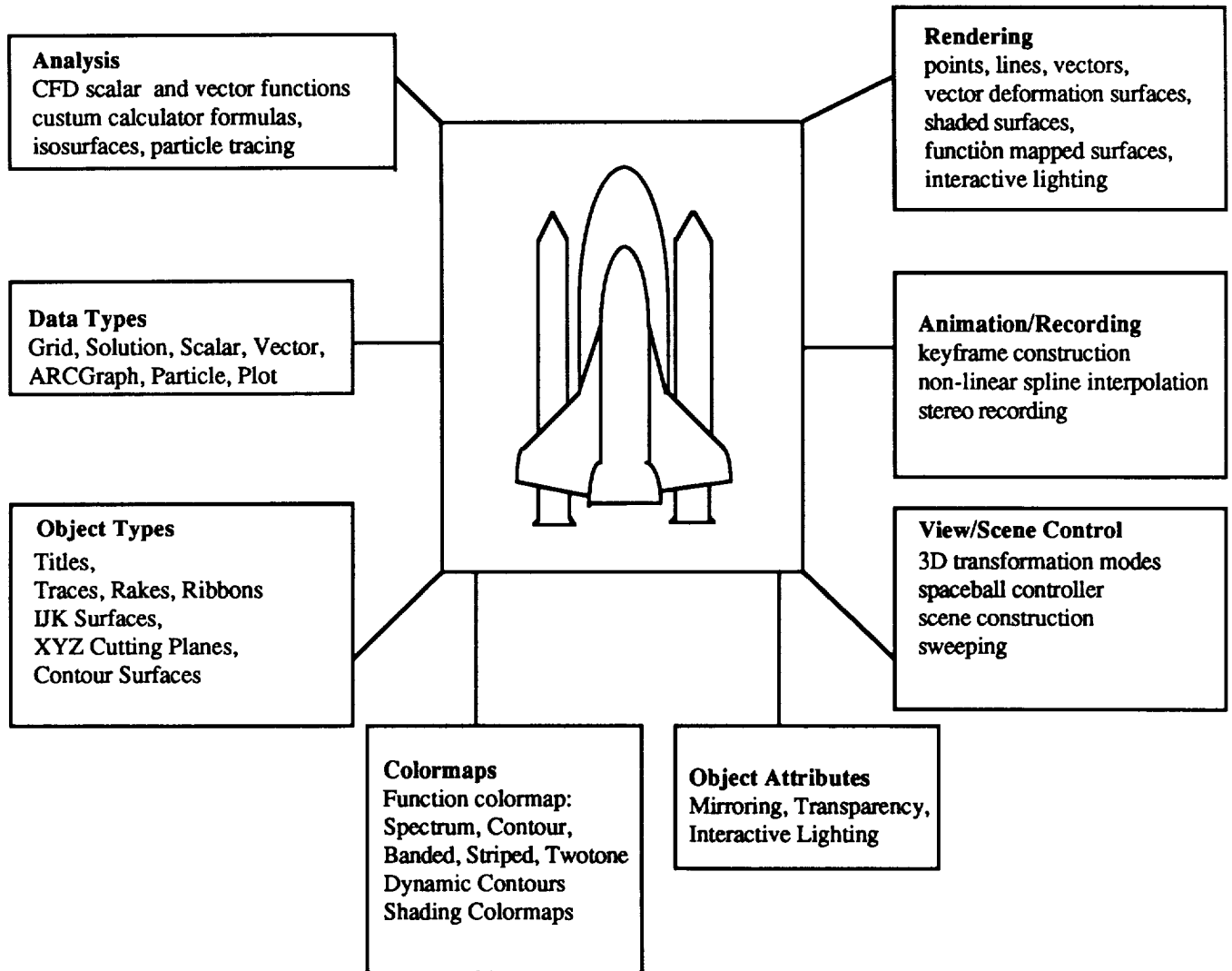


Figure 1, "FAST Interactive Visualization Control"

### Interactive Visualization Control

**Multi-processed:** When working within the FAST environment, several modules work together to generate a scene: Surfer generates the grid surface objects, CFD Calculator computes the scalar and vector fields, Titler is used to generate the text, and FAST Central was used for image handling and color map editing. When modules are not needed they can be iconified so they occupy less screen space and CPU resources. Because of this, the FAST environment can be running while other applications are also being used. Alternatively, FAST modules can be terminated without exiting the FAST environment - and this has no effect on their data since it is already in shared memory. Unlike standard dynamic memory, shared memory remains available even after the allocating process is killed. All shared memory segments are removed when FAST is exited via the Quit selection of the FAST Central module.

**Powerful.** The FAST environment contains sophisticated tools such as the CFD Calculator that enable the scientist to analyze computed solutions by examining many relevant "CFD quantities", such as normalized helicity, shock, perturbation velocity, and vorticity. And if these "built-in" functions are not adequate the

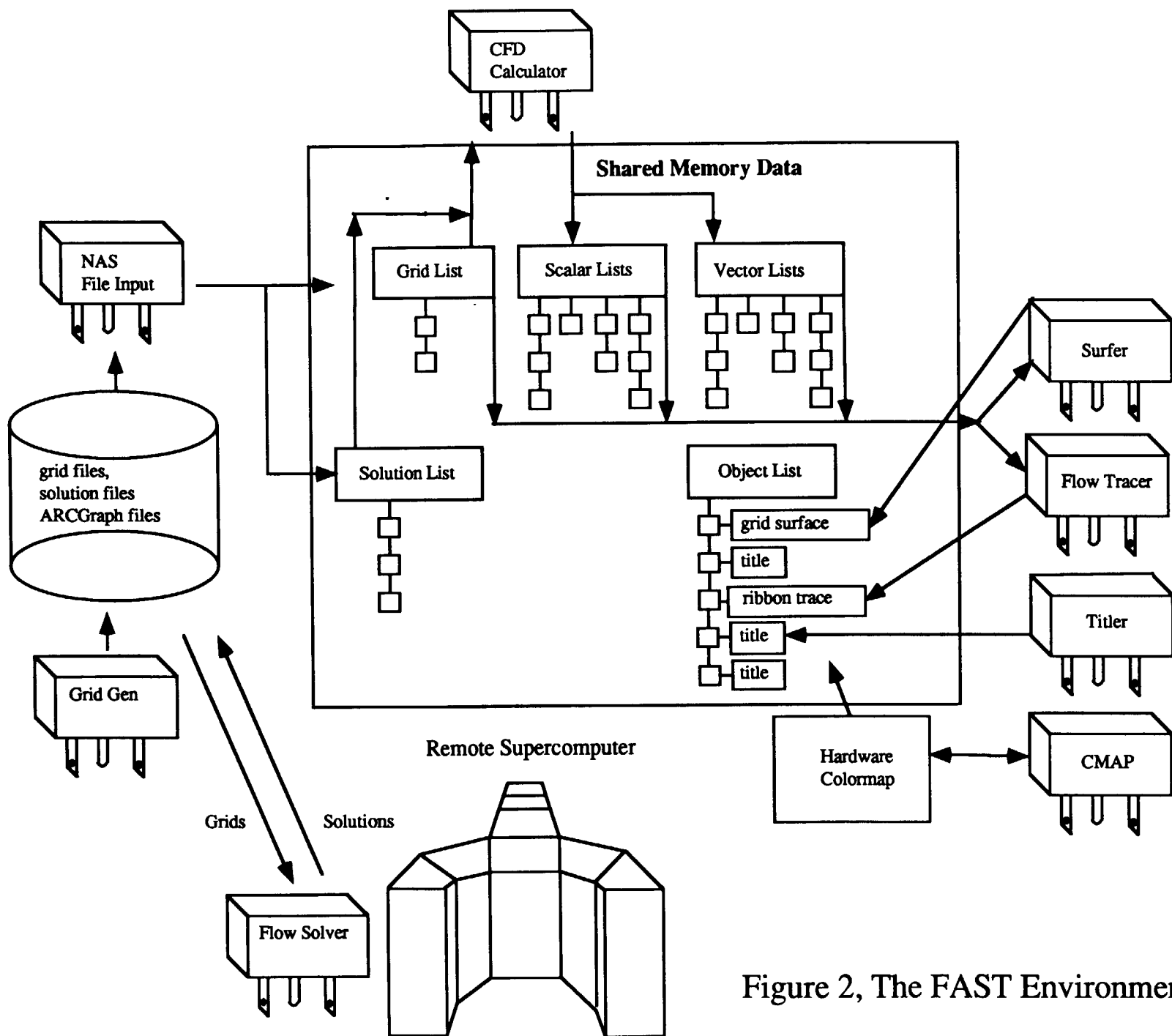


Figure 2, The FAST Environment

scientist can program the Calculator to compute customized functions using the rich set of component, scalar, and vector operators.

**Flexible.** Storing data and graphical objects in shared memory enables complex scenes to be constructed by mixing and matching shared data from any module that is currently plugged into the FAST environment. This allows grid, scalar, and vector data to be combined to generate grid surfaces rendered as grid lines, scalar colored smooth polygon surfaces, and vectors.

**Interactive.** Surfer provides the ability to interactively alter scene attributes such as coloring the data by a different scalar field, displaying a different vector field, adjusting the legend, normalization, and clipping ranges, or changing rendering and data types. For example, the vector field can be rendered as a Gouraud shaded, lighted, vector deformation surface. With the looping option turned on Surfer will sweep through all data in the current grid direction - providing a dynamic visualization ability. And while this is happening the scientist (from FAST Central) can transform (e.g., rotate, or zoom) all or part of the scene or use the color map editor to adjust the function color mapping by inserting, deleting, and changing colors, or selecting a different colormap types such as *Spectrum*, *Contour*, *Striped*, *Twotone*, or *Banded*.

### IPC and Shared Memory Implementation

It was decided that an interprocess communication (IPC) package must be implemented to allow FAST to operate as a modular environment where resources could be shared among different machines as well as on a single host. Specifically, Unix System V shared memory facilities are used to allow each process (module) to access the environment's data, while the Berkeley IPC package's implementation of Internet domain stream sockets allows for the coordination of this data.

As each module is executed by the FAST hub, it must immediately establish a two-way communication channel between it and the hub. Because an Internet domain address consists of a machine network address and a port number, these two values are used in establishing this connection. The following command is therefore executed at the beginning of a module's main routine:

```
socket_establish_and_accept (hub_host, hub_port, &rsock, &wsock);
```

This does the following:

- 1) create a socket from which to read
- 2) determine a local port and listen on it
- 3) create a writeable socket and establish a connection to the hub (using the hub's hostname and port number which came in as arguments)
- 4) now send the port number to the hub and
- 5) accept a connection from the hub

At the same time, the hub process executes this statement:

```
socket_accept_and_establish( sock, module_host, &wsock );
```

which does the following:

- 1) create one socket from which to read from all modules
- 2) accept a connection from the next module
- 3) read in the module's port number
- 4) create a writable socket and connect this socket to the module

The modules specified for inclusion in the FAST environment are specifically listed in a "run command" file called \$HOME/.fastrc. Also included within this file is information about initial placement of a module's main panels, the name of the host where the module resides, and the complete path name of the particular module.

Once a module has been executed by the FAST hub using the Unix system(3) call and the communication channels have been established, the hub enters a loop where it waits on a request from any of the active



modules to perform some sort of action. The hub process uses the Unix `select(2)` call to examine all available read socket file descriptors to determine if there is information in the pipe.

Information sent between a module and the hub (and vice-versa) is always preceded by a standard sized structure which contains, the command and four information fields. The necessary information, if any, is then written back to the module, and the flow control takes the hub back up to the point where it can again wait for a request.

One example of a request that a module might make is the allocation of memory which can eventually be used by another module. It must first send a request to the hub to do this. The hub then allocates the memory as a shared memory segment and retrieves the shared memory identifier associated with this segment. This identifier is then stored by the hub in a data structure possibly to be accessed by another module at a later time. Finally this identifier is sent back to the module so that it may attach the shared data to its virtual memory address space.

At any time that a different module would like to access this data, a request is similarly sent to the hub to retrieve the shared memory identifiers so that it too may attach to the data. Once the module has attached to the data, UNIX semaphores are implemented to lock the shared data if it must be modified.

A consequence of using shared memory instead of standard dynamic memory is that dynamic data structures such as linked list nodes no longer have a *pointer* to the next node but rather the *shared memory id* of the next (and current) node. This shared memory id must be explicitly *attached* to and *detached* from whenever the structure is traversed.

Using shared memory and sockets, FAST is able to quickly and easily share all the data used within the environment. Even though shared memory can not yet be shared over different machines as it is on a single host, FAST has been designed with this eventual capability in mind. When indeed we can accomplish this, the ultimate power of FAST can be realized.

## **Discussion**

For an existing SGI visualization application to be converted into a FAST module:

- Command line arguments must be used to establish window location and Hub communication - and nothing else.
- Periodically, each module must check for a command from the Hub. This is done once each time through the main event loop.
- Standard input should not be used.
- Standard output should be used sparingly for status and error messages.
- The colormap must be used according to FAST conventions. FAST library functions must be used to get color indices for drawing. A few indices are reserved for modules to create their own colors, but most of the colormap is only modified via the FAST COLORMAP module.
- Grid, vector and scalar field data must be accessed via FAST shared memory.
- The panel library interface should be used for module interaction.
- The panel library's nap time or blocking should be turned on when waiting for user input to avoid excessive context switching.
- The application's drawing code must be integrated into the viewing library so that its visualizations can appear in FAST central.
- The data to be displayed must be placed in shared memory and made available to the viewing mode

There are several advantages to integrating applications into FAST as modules. These advantages include:

- Shared memory speeds which allow users to interactively view their data from several modules without long disk IO delays.
- Access to CFD Calculator to generate vector and scalar fields.
- Precalculated min and max for grids, vector and scalar fields. This reduces the time needed to access data in many cases.
- Sophisticated colormap manipulation using the FAST colormap panel.
- Integration of visualizations created by several modules into a single scene.

- Trivial integration of visualizations into animations.
- New applications can be built quickly since many functions have already been made available in existing FAST modules and libraries.

There are also some disadvantages, of course. These include:

- Time to learn to use the FAST libraries and intermodule communications as well as to keep up with future changes.
- Performance overhead due to multiple processes busy waiting.

Future plans for FAST include the capability for use across high speed LANs for 'smart' distribution of processing. Compute intensive modules could be distributed or broken up into components that communicate over these networks, or perhaps memory could be shared across systems.

As flow solvers become fully integrated and interactive 3-d grid generation becomes a reality, FAST will continue to offer more effective visualizations of computational aerodynamics in all aspects of fluid flow simulations.

At the time of this writing, the software is in Beta testing at NASA Ames Research Center and at over fifty sites around the country. FAST 1.0 is tentatively scheduled for release in late 1990. Since this software was developed for the government, it is in the public domain, and is available for no charge or a minimal handling fee.

## References

1. P. Buning et al., "Flow Visualization of CFD Using Graphics Workstations", AIAA 87-1180, *Proc. 8th Computational Fluid Dynamics Conf.*, June 9-11, 1987.
2. J. Foley and A. Van Dam, *Fundamentals of Interactive Computer Graphics*, Addison Wesley, 1982
3. P. Haeberli, "ConMan: A Visual Programming Language for Interactive Graphics," SIGGRAPH Proceedings, Volume 22, Number 4, SIGGRAPH August 1988
4. C. Upson, et al., "The Application Visualization System (AVS): A Computational Environment for Scientific Visualization" IEEE Computer Graphics and Applications, July 1989
5. J. Helman, L. Hesselink, "Representation and Display of Vector Field Topology in Fluid Flow Data Sets", IEEE Computer, August 1989
6. G. Bancroft, "Scientific Visualization in Computational Aerodynamics at NASA Ames Research Center", IEEE Computer, August 1989
7. P. Walatka, P. Buning, *PLOT3D Users Manual Version 3.6* NASA TM101067, 1989
8. T. Plessel, *SURF Users Manual*, NASA Ames Research Center, Code RFW, 1988
9. T. Plessel, *GAS Users Manual*, NASA Ames Research Center, Code RFW, 1988
10. S. Leffler, et al., "The Design and Implementation of the 4.3 BSD UNIX Operating System", Addison-Wesley, 1989
11. K. Haviland, B. Salama, "UNIX System Programming", Addison-Wesley, 1987
12. *Stellix Programmers Guide*, Chapters 15-17, Stellar Computer Inc., 1988
13. Lorenson, W.E., and Cline, H.E., "Marching Cubes: a High Resolution 3D Surface Construction Algorithm," Computer Graphics, Vol 2.1, No. 4 July 1987, pp. 163-169.
14. D. Tristram, P. Walatka, E. Raible "Panel Library Programmers Manual, Version 9.5", NASA Ames Report RNR-90-006
15. F. White, "Fluid Mechanics", McGraw-Hill 1979
16. D. Anderson, "Computational Fluid Mechanics and Heat Transfer", McGraw-Hill 1984
17. L.M. Milne, "Theoretical Aerodynamics", Dover Press 1973
18. T. Lasinski, "Second Generation Graphics Workstations, Request For Proposals", RFP2-33491, January 1989
19. S. Rogers, "Distributed Interactive Graphics Applications in Computational Fluid Dynamics", International Journal of Supercomputing Applications, Vol 1, No. 4, Winter 1987
20. E. Hibbard, "ARCGRAPH (Ames Research Center Graphics Metafile) Manual", NASA Ames Research Center

## HYPERCUBE TECHNOLOGY

Jay W. Parker, Tom Cwik, Robert D. Ferraro, Paulett C. Liewer, Jean E. Patterson

Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California 91109

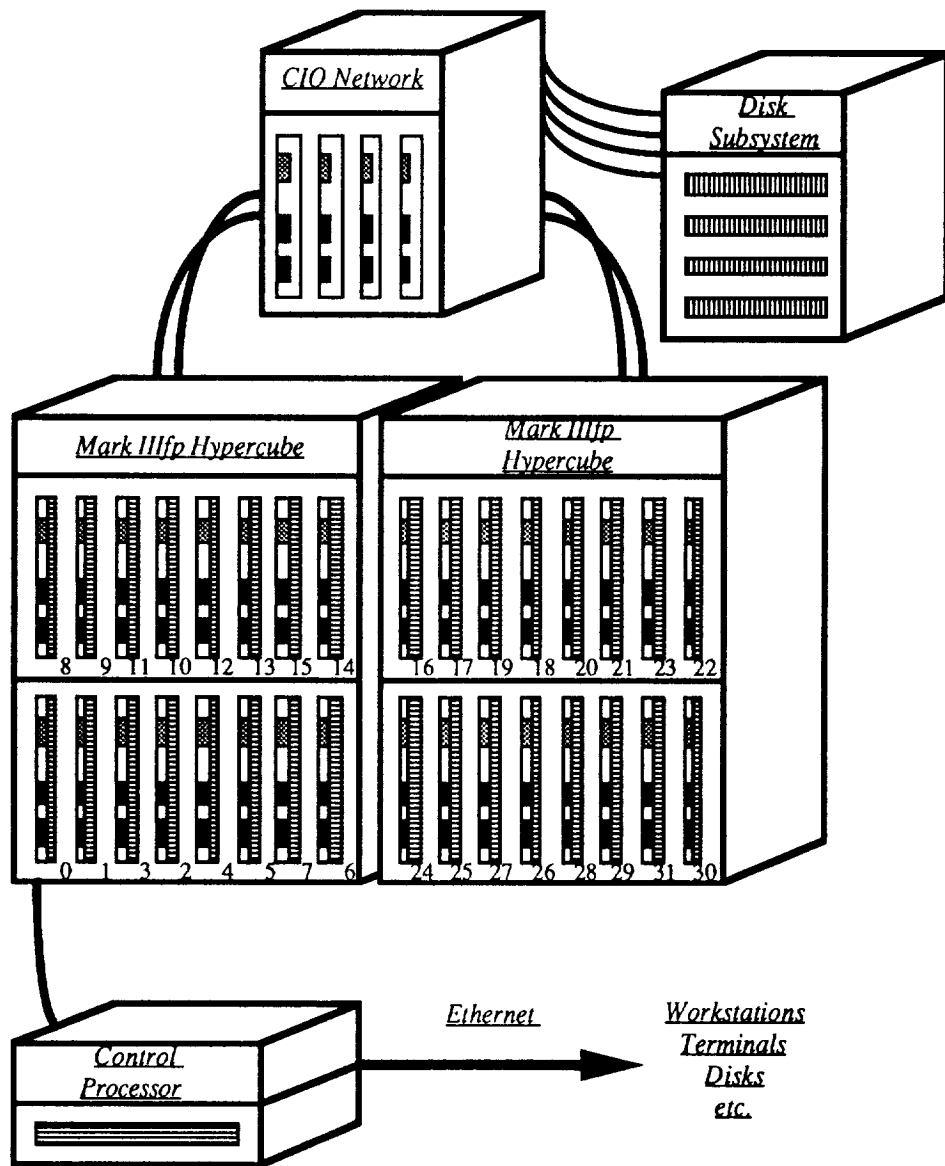
**Abstract:** The JPL-designed MARKIII hypercube supercomputer has been in application service since June 1988 and has had successful application to a broad problem set including electromagnetic scattering, discrete event simulation, plasma transport, matrix algorithms, neural network simulation, image processing and graphics. Currently, problems that are not homogeneous are being attempted, and, through this involvement with "real world" applications, the software is evolving to handle the heterogeneous class problems efficiently.

**Hypercube Computers and the Mark IIIfp System:** The appearance of parallel computers comes through a natural coupling of two important factors. The first one is the clear *opportunity* presented by current VLSI technology, which determines the availability of powerful single-chip microprocessors and inexpensive memory. The second one is the pressing *need* for increased computational power in a wide variety of scientific and engineering fields. A particularly simple parallel computer consists of loosely coupled processors connected in a binary hypercube topology. A  $D$ -dimensional hypercube contains a total of  $D2^{D-1}$  communication links, and  $2^D$  nodes where each node is connected to  $D$  other nodes. One of the most useful characteristics of this topology is that the diameter of the network (the maximum number of links that a message has to travel between any source and any destination along the shortest path) is also equal to  $D$ . The JPL/Caltech Hypercube is now in its third generation of development. The current implementation in the series has been designated the Mark IIIfp Hypercube, and is based on Motorola and Weitek chip sets, as described below. A description of the earlier Mark III and several related parallel computers, their software environment, and sample applications approaches and codes may be found in [1]. Further details regarding parallel operating systems and applications may be found in [2].

Fig. 1 shows the general system description of a 32-node (5-dimensional) Mark IIIfp Hypercube system. The main components of the system are shown: a 32-node Mark IIIfp, the control processor, the Concurrent I/O (CIO) network and the disk subsystem. Each node in the Mark IIIfp is a powerful single-board computer which contains two independent processing elements, namely the *data processor* and the *I/O processor*. Each node contains its own local memory and a set of peripherals. In addition, each node contains 4 Megabytes of dynamic RAM accessible to both processors. Each node has a total of eight communication channels, one of which is reserved for communications outside the hypercube. Thus, the current architecture is limited to 128 nodes, or a 7-dimensional Mark IIIfp.

The data processor consists of a Motorola MC68020 CPU with a MC68882 floating point math coprocessor, two serial ports, and a printer port. The serial ports can be used to connect a terminal to the node to monitor the activity of the data processor. Associated with the data processor are 128K bytes of private no-wait-state static RAM. Programs run approximately 15% faster when stored within this memory space. The I/O processor consists of a Motorola MC68020 CPU, one serial port, and hardware to support the node-to-node communications within the hypercube. The serial port can also be used to connect a terminal to the node and monitor the activity of the I/O processor, independent of the data processor. Associated with the I/O processor are 64K bytes of private no-wait-state static RAM.

The floating point daughterboard incorporates a Weitek chip set and a serial port (the letters "fp" on the designation of the Mark IIIfp hypercube denote the active inclusion of the Weitek daughterboard into the architecture). The Weitek chip set consists of a sequencer, an integer processor, and a pipelined floating point processor, and is capable of performing from 1 to 10 MFLOPS. Typical applications written in C or FORTRAN with calls to CrOS communications achieve the lower end of this range. Many such applications maintain near-linear performance when scaled to larger hypercubes; one can expect real application performance for a 128 node hypercube to be on the order of 100-300 MFLOPS.



**Figure 1.** A 32-node (5-dimensional) JPL/Caltech Mark IIIfp Hypercube system. This figure shows the main components of the system: the hypercube, the control processor, the CIO network and the disk subsystem. The connections between these components are also shown, including the interface to the outside world through Ethernet.

Recently, the Weitek board has been upgraded to support 64-bit double-precision arithmetic. The serial port on the Weitek daughterboard can be used to connect a terminal to monitor the activity of the Weitek processor, independent of both the data processor and the I/O processor. The Weitek processor has access to its own private memory, which consists of 64K bytes of no-wait-state static RAM as well as a 128K byte code cache and a 64K byte data cache.

The Mark IIIfp is accessed from the external world through a host or control processor (CP) computer, which in turn communicates with the hypercube through a JPL-designed, special-purpose interface connected to the corner node (node 0) of the ensemble. The control processor is a Unix workstation based on the Motorola MC68020 microprocessor and currently is a Counterpoint System 19. The CP can be attached to peripheral devices such as disk drives, terminals, and printers and acts as an access controller mechanism to these devices for the entire hypercube. Since the control processor is based on the same

MC68020 microprocessor used in the node boards, the native compilers and linkers, as well as the full complement of Unix tools can be used for program development. These tools allow the construction of executable code on the control processor which ultimately runs on the nodes of the Mark IIIfp.

A concurrent I/O (CIO) node board resembles a Mark IIIfp's main board but has a single MC68020 processor and only four communication channels. In addition to the ability to form a hypercube network consisting of CIO nodes, a CIO board has a VME bus interface which allows the Mark IIIfp to communicate with disks, graphics devices or specialized processing boards. A single CIO board can be connected to up to 4 nodes within the hypercube. This allows different Mark IIIfp nodes to simultaneously place requests for accessing devices connected to a CIO board through its VME bus interface.

The operating system on the control processor is AT&T's Unix System V, Version 2.2. Routines have been provided to allow the interface between the control processor and programs running on the nodes of the Mark IIIfp under the control of the CrOS or Mercury operating systems described below. Detailed reference information suitable for Mark IIIfp applications development utilizing these operating systems is given in [3].

The Crystalline Operating System (CrOS) provides the programmer with the fastest possible communications scheme and is well suited for applications where the domain of the problem can be correctly mapped onto the hypercube topology. The most important characteristics of CrOS are:

- Communication is done between adjacent nodes only. Data transfers are only supported among "nearest neighbor" processing elements which are physically connected to one another.
- Communication is channel based. Each node designates a particular *channel number* to each of the links to its neighboring nodes in each dimension. CrOS communication routines make use of this channel number as their main parameter.
- Communication is synchronous. A request for exchange of data by one node must be matched by a corresponding request in the neighboring node, otherwise, a system deadlock would occur.
- High transfer rate. The bandwidth of this style of communication is 2 Megabytes per channel per second, with each node being capable of communicating data through all its communication channels simultaneously.

The Mercury operating system is used to allow asynchronous message passing between nodes. In addition, it is recognized that certain applications require asynchronous services, but which also contain certain portions which operate in a tightly coupled way and can benefit from the higher transfer rates available through the use of synchronous communication. In order to accommodate to these situations, the Centaur operating system was built on top of both CrOS and Mercury. Centaur gives the programmer extra flexibility by allowing the application program to switch between synchronous communication mode making use of CrOS services and asynchronous communication mode making use of Mercury services.

Support is provided for one user per Mark IIIfp hypercube, with a maximum of one main program running on each node of the system and a different program running on the control processor. Application programs running under CrOS may be written in C or FORTRAN. Those running under Mercury can currently be written in C only.

**Applications:** Many types of applications have been written or ported from sequential machines to run on the Mark IIIfp hypercube. Several are described in [2]. The following is a partial list:

- Electromagnetics Scattering Analysis:
  - Finite Difference
  - Methods of Moments
  - Finite Element

- Concurrent Database Implementations
- Image Processing
- Geophysical Modeling
- Lattice Gauge Simulations
- Computer Chess and Game Theory
- Vortex Flow Simulations
- Optimization Problems
- Lisp Interpreter
- Distributed Prolog
- Neural Network Simulations
- Neural Network Applications:
  - Early Vision
  - Pattern Recognition
  - Optimization
- Tracking Algorithms
- Battle Management Simulations
- Studies in Plasma Turbulence
- Plasma Particle Simulations
- Discrete Event Simulations
- Studies in Radiative Transfer
- Ray Tracing Algorithms

**Electromagnetic Scattering by Finite Elements:** In order to illustrate some of the issues involved in developing and running applications on a hypercube computer, we shall examine in detail an electromagnetic scattering code, which employs the technique of finite elements to obtain solutions. Early stages of this work, and several other parallel codes which solve scattering and related electromagnetic problems employing the method of moments, time-domain finite differences, and methods for frequency selective surfaces, are described in [4].

The well-posed scattering problem consists of a set of scatterers composed of dielectric and conducting materials with possibly anisotropic and inhomogeneous dielectric properties. A known electromagnetic wave illuminates the objects. The electromagnetic field scattered from the objects is to be calculated at infinity (the far field), and possibly close to the scatterers (the near field) as well. Since the finite element method requires a meshed problem space, some outer boundary condition is required which will impose an outgoing wave solution on the scattered field. Because of computational limitations, the outer boundary needs to be as close as possible to the scatterers. The wave equation for either the electric or magnetic field is solved in the frequency domain subject to the outer boundary condition, the boundary conditions imposed by the presence of conductors, and any simplifications allowed by the geometry.

The finite element mesh (nodes and elements) is generated to express the geometry of the scatterers, including their internal constituents. The node density and element types used to model the problems are chosen based on the specific accuracy required in the solution and the limits imposed by computational resources. The issue of mesh generation is a complicated one in itself. The density of nodal points and the complexity of element type will impact the accuracy of the solution and the computation required to compute that solution. The complexity of the outgoing wave boundary condition imposed at the outer problem boundary, as well as the position, shape, and grid density of the outer boundary also impact the solution accuracy.

We take as our basic equation the Helmholtz equation for either the electric or magnetic field in linear media, in the absence of free charges. The equation for  $\mathbf{E}$  is given by

$$\nabla \times (\mu^{-1} \cdot \nabla \times \mathbf{E}) - \frac{\omega^2}{c^2} \epsilon \cdot \mathbf{E} = 0 \quad (1)$$

where  $\epsilon$  is a general electric permittivity tensor,  $\mu$  is a general magnetic permeability tensor,  $\omega$  is the angular frequency of the field, and  $c$  is the speed of light. The equation for  $\mathbf{H}$  is obtained by simply

swapping  $\epsilon$  and  $\mu$  throughout. In general, the dielectric functions may be complex, to model absorptive media. Either the  $\mathbf{E}$  or  $\mathbf{H}$  equation with boundary conditions is a sufficient statement of the problem.

The weak form equation required to employ the finite element method is obtained by taking the inner product of Eq. 1 with a test function  $\mathbf{T}$  and integrating over the problem domain. The test function must satisfy some basic constraints on continuity and integrability, but is otherwise arbitrary. Thus the weak or Galerkin form equation for the electric field is

$$\int_{\Omega} d^3x \left[ (\nabla \times \mathbf{T}) \cdot \mu^{-1} \cdot (\nabla \times \mathbf{E}) - \frac{\omega^2}{c^2} \mathbf{T} \cdot \epsilon \cdot \mathbf{E} \right] = \int_{\Gamma} d^2x (\mathbf{n} \times \mathbf{T}) \cdot \mu^{-1} \cdot (\nabla \times \mathbf{E}) \quad (2)$$

where  $\Gamma$  is the contour which bounds the problem domain  $\Omega$ . A similar equation may be obtained for  $\mathbf{H}$ .

In order to find the electric field solving Eq. 2,  $\mathbf{E}$  and  $\mathbf{T}$  must be restricted to a linear space of finite dimension. In the finite element method, this linear space is a finite summation over a set of basis functions  $\mathbf{w}_n$ ,

$$\mathbf{E} = \sum_n d_n \mathbf{w}_n \quad (3)$$

$$\mathbf{T} = \sum_n c_n \mathbf{w}_n \quad (4)$$

each of which is nonzero only in a small region of space. Interpolatory basis functions are preferred for most problems; these bases are defined to be unity at a single node of the mesh, and zero at all other nodes, typically based on simple polynomials defined in each polyhedral region of the mesh. The advantage of interpolatory basis functions is that each weight  $d_n$  in the summation for  $\mathbf{E}$  may be interpreted directly as the value of the solution at the corresponding node.

The assumption linking Eq. 1 and Eq. 2 is that the field  $\mathbf{E}$  which satisfies Eq. 2 for *any* test function  $\mathbf{T}$  (arbitrary within the space of functions satisfying certain criteria of continuity and boundary conditions) must be precisely that  $\mathbf{E}$  which solves Eq. 1. Substituting Eqs. 3 and 4 into Eq. 2 results in

$$\mathbf{c} \cdot \mathbf{K} \mathbf{d} = \mathbf{c} \cdot \mathbf{f}, \quad (5)$$

and the arbitrariness of  $\mathbf{T}$  implies that this must hold for arbitrary  $\mathbf{c}$ , therefore the  $\mathbf{d}$  must satisfy

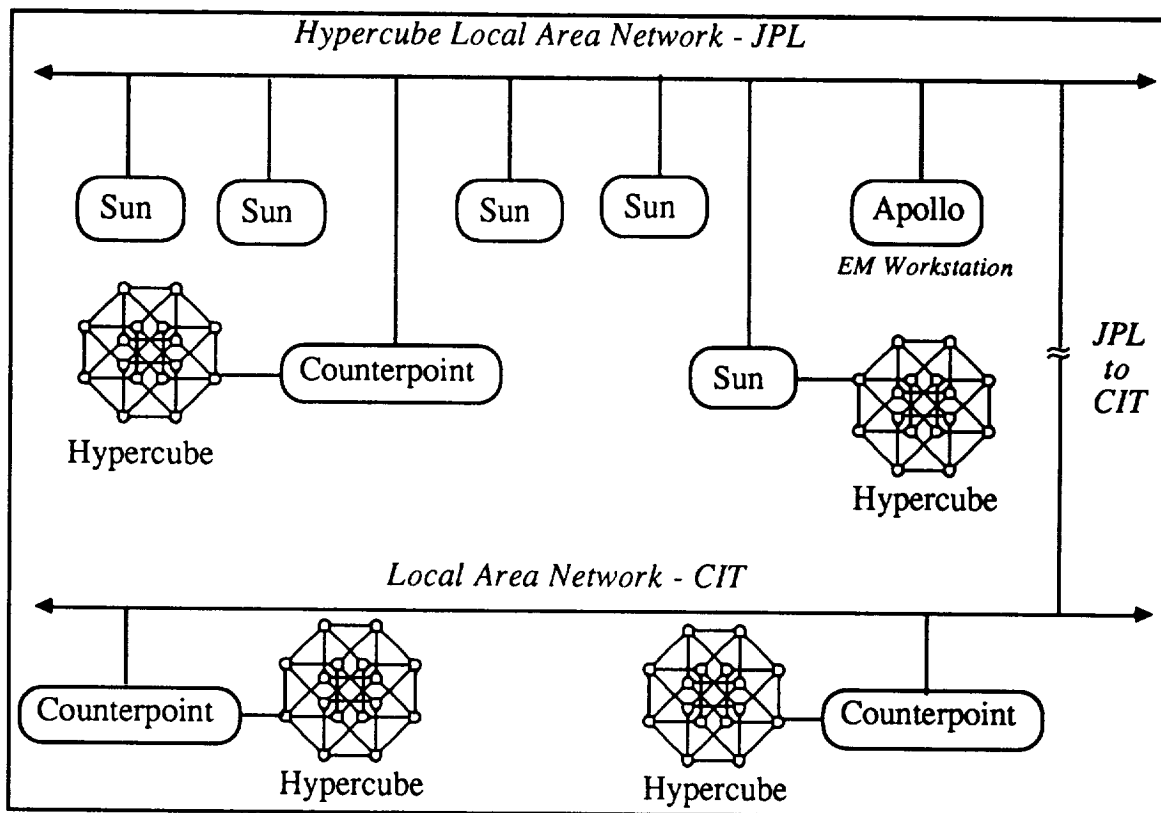
$$\mathbf{K} \mathbf{d} = \mathbf{f}. \quad (6)$$

This system of equations is inverted to find  $\mathbf{d}$ , which determines the field everywhere in the computational domain according to Eq. 3. The matrix  $\mathbf{K}$  consists of integrals of the basis functions combined in pairs, with typical elements of the form

$$k_{ij} = \int_{\Omega} d^3x \left[ (\nabla \times \mathbf{w}_i) \cdot \mu^{-1} \cdot (\nabla \times \mathbf{w}_j) - k_0 \mathbf{w}_i \cdot \epsilon \cdot \mathbf{w}_j \right] \quad (7)$$

Additional terms are required to implement certain boundary conditions, such as those at conductors, material interfaces and the computational domain truncated boundary. Eq. 7 (and the additional terms) imply several computationally important features of the system of equations to be solved: the matrix is symmetric, and extremely sparse, containing non-zero entries only when two basis functions share some portion of their non-zero domain. This sparsity, in combination with an appropriate parallel sparse equation solver on a hypercube computer, allows solution of problems with several hundred thousand unknowns.

**Workstation Environment:** In order to rapidly take advantage of the vast computational power of the Mark IIIfp supercomputer in an environment with a growing number of electromagnetic analysis parallel programs, we have developed an Electromagnetics Interactive Workstation, based on an Apollo DM4500 graphics computer connected to various Mark IIIfp hypercubes by a high-speed network. Some of the key features of the EIAW include a consistent and friendly interface based on a modular software design, transparent access to remote computing resources and an inherent ability for further expansion. The user is initially presented with a menu, enabling quick choices of graphical tools for model building and results display, interactive selection of parameters for a computation, a variety of electromagnetics codes, and an assortment of sequential or parallel computers which may be either local or remote. Fig. 2 shows the current location of the EIAW within our local area network. The network allows direct access to Sun, Counterpoint and Apollo machines as well as indirect access to several Mark IIIfp Hypercube systems. Program development is mostly done on Sun workstations, with executables being sent over the network to Counterpoint systems which currently act as host machines for the existent Mark IIIfp Hypercubes.



**Figure 2.** A sketch of the physical arrangement of the resources available through the JPL and CIT LANs.

Given a developed code, a design engineer may access the appropriate software and hardware resources for a given electromagnetic scattering problem from the workstation interface without needing to know anything about the network or the parallel computers. In a typical session, the engineer would start up the EIAW menu, and by simple mouse/cursor selection move through the design and testing of the scattering object as follows. First, the analysis tool, Finite Elements, is selected from a submenu from among such choices as FDTD (a Finite Difference Time Domain code) and Patch (a method of moments code). Next, a simple name for the object is chosen. The system will use this name, with various extensions, to create file names with a simple and consistent naming convention. Via another submenu, the engineer selects a model-building tool, which may be a commercially available CAD program. After the model is constructed, the model-building tool returns the engineer to the EIAW menu. Another submenu choice allows the engineer to enter model parameters such as the incident field description. The engineer then selects which computation engine to use to run the analysis program. At another menu command, the system partitions



the model into the appropriate number of pieces for the parallel computation, sends the model to the computation engine, runs the analysis program, and returns the solution field to the local environment. Post-processing utilities may then be selected to locally compute or display derived quantities of interest, such as the radar cross section of the object.

**The Parallel Algorithms:** The finite element method readily lends itself to parallel execution. The chief parts of the computational burden are the creation of the  $\mathbf{K}$  matrix, and the solution of the sparse system of equations. The form of the integrand in Eq. 7 implies that non-zero entries in  $\mathbf{K}$  arise from node couplings through polyhedral regions (elements) which contain a given pair of nodes. Each element couples all possible pairs of the nodes it contains, contributing entries in  $\mathbf{K}$  which correspond to these node pairs. Because nodes are typically shared by several elements, a given node is coupled to all of the nodes in all of the elements of which it is a part, and only to those near-neighbor nodes. Therefore, a natural way to divide the problem is to divide the spatial domain. We give groups of neighboring elements and the nodes which belong to them to a single processor. Some nodes must be shared, which implies communication among the processors; thus it is advantageous to minimize the number of these shared nodes. In order to finish in the shortest time, each processor must do approximately an equal share of the work, and must therefore have an equal share of the model.

With this in mind, we have developed software to divide a model automatically and nearly optimally. We use an algorithm called Recursive Inertial Partitioning. The essential idea is to find the long axis of the model, and divide midway with a plane perpendicular to this axis. A moment computation is performed, corresponding to calculating the moment of inertia tensor, with each element having unit weight at its center. A plane perpendicular to the largest moment slices the object through the center of mass, with whole elements assigned to one side or the other. Each section is then redivided according to the same algorithm. Thus the object is divided into  $2^D$  groups of elements, ready to be processed on a  $D$ -dimensional hypercube. Fig. 3 shows a sample finite element domain as divided by the RIP algorithm.

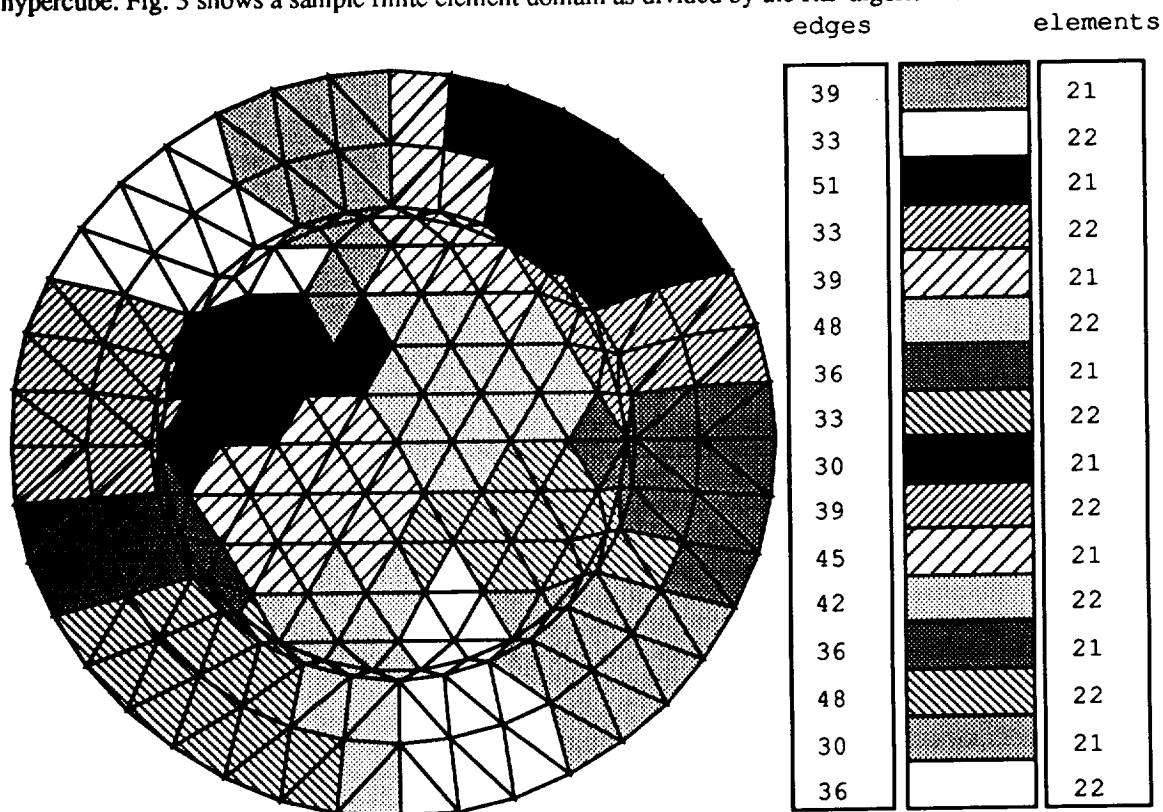


Figure 3 - Decomposition of a grid describing a cylindrical dielectric scatterer, using the RIP method implemented in the SLICE partition code. Load balance of elements and boundaries are illustrated.

Each processor of the hypercube computes its share of the  $K$  matrix completely independently, that is, with no interprocessor communication. This is achieved by redundantly storing the information for shared nodes in both processors, and by allowing the shared portions of the  $K$  matrix to remain unsummed. That is, when a  $K$  matrix entry involves two nodes each shared by two processors, that entry consists of a sum of terms, some of which are contributed by elements in the first processor, and some contributed by elements in the second. There is no need to explicitly sum these terms into one memory location, because the solution portion of the code can find all the appropriate partial sums when they are needed.

Solution is performed by a Preconditioned Bi-Conjugate Gradient algorithm. This is an iterative algorithm, with the drawback that the convergence to acceptable accuracy is not entirely predictable as to the number of steps. Convergence would occur with no error after  $N$  steps for a rank  $N$  system of equations, if the algorithm were performed with infinite precision arithmetic. In practice, well-posed problems typically converge to acceptable accuracy in less than  $N/10$  steps. The great advantages of the method are that very little communication is required between the processors, and all steps of the algorithm can be done using the memory locations describing the system of equations. That is, in contrast to a banded-matrix solver, or (even better) a Crout decomposition, no matrix locations which are initially zero are filled in with non-zero values at any stage of the algorithm. Therefore, we store the non-zero matrix entries as an indexed list of numbers; we need never store zero entries.

The computationally significant parts of the PBCG algorithm are vector dot products and matrix-vector multiplies. Given distributed data storage as described above, two peculiar features of these operations as done in parallel are noted here. First, as to performing dot products, some vector components (corresponding to nodes in the finite element model which are shared among several processors) are replicated in several processors, while the contribution to the dot product from these components must be computed only once. Therefore, although these nodes are shared, only one processor has these nodes flagged as owned by that processor (the flagging having been done at the time of model partitioning). Partial dot products in each processor are computed using only vector components corresponding to owned nodes. Then the partial sums are globally added, a step requiring interprocessor communication. Second, with regard to matrix-vector multiplies, recall that  $K$  matrix entries corresponding to shared nodes consist of a sum of terms, and that these terms have not been explicitly summed, but rather reside in separate processors. The correct result may be obtained by having each processor do a matrix-vector multiply with respect to only the portion of the matrix and vector which reside in that processor, and then summing the results in an appropriate way with each processor contributing its portion (again, a step requiring interprocessor communication). In both of these operations, the amount of interprocessor communication is small compared to the amount of computation within each processor, provided the model consists of a large number of elements and nodes *per processor*.

**Results:** In order to demonstrate the type of object which may be modelled with the 2-D electromagnetic scattering code, and the sort of accuracy one may obtain easily, we compute the near and far field for a perfectly conducting circular cylinder with radius such that  $ka = 50$  (i. e., the circumference is 50 wavelengths). The field is shown in Fig. 4. Since this is an object of simple geometry, one may also compute the solution analytically, and a comparison of the two solutions demonstrates the accuracy of the finite element solution. As seen in Fig. 5, good agreement (within 3 dB over a 30 dB range) with the analytic RCS was obtained by truncating the computational domain at  $kr = 62$ , while  $kr = 56$  was not considered adequately accurate (errors exceed 4 dB). The more-accurate case involved less than 10,000 unknowns, which is a fairly small problem for the Mark IIIsp. With a 64 node Mark III, we can solve similar problems with over 200,000 unknowns, providing the ability to achieve either much higher accuracy, or to model larger and more complex objects.

The time to realistically solve a given problem depends on I/O rates as well as the time to set up and solve the finite element system of equations. Since I/O is currently done sequentially, there is a problematic bottleneck involved in reading the model and writing the fields from the hypercube. This I/O cost may be amortized by solving a large number of scattering problems based on the same model, such as by varying the incident field angle, the wavelength, or the properties of materials in the object. However, the I/O bottleneck can be significant when only one problem solution is needed, and we are investigating means of ameliorating the problem, such as generating the finite element mesh in parallel on the hypercube. Apart from I/O, the time required to solve a finite element problem is based on the time to set up the system of

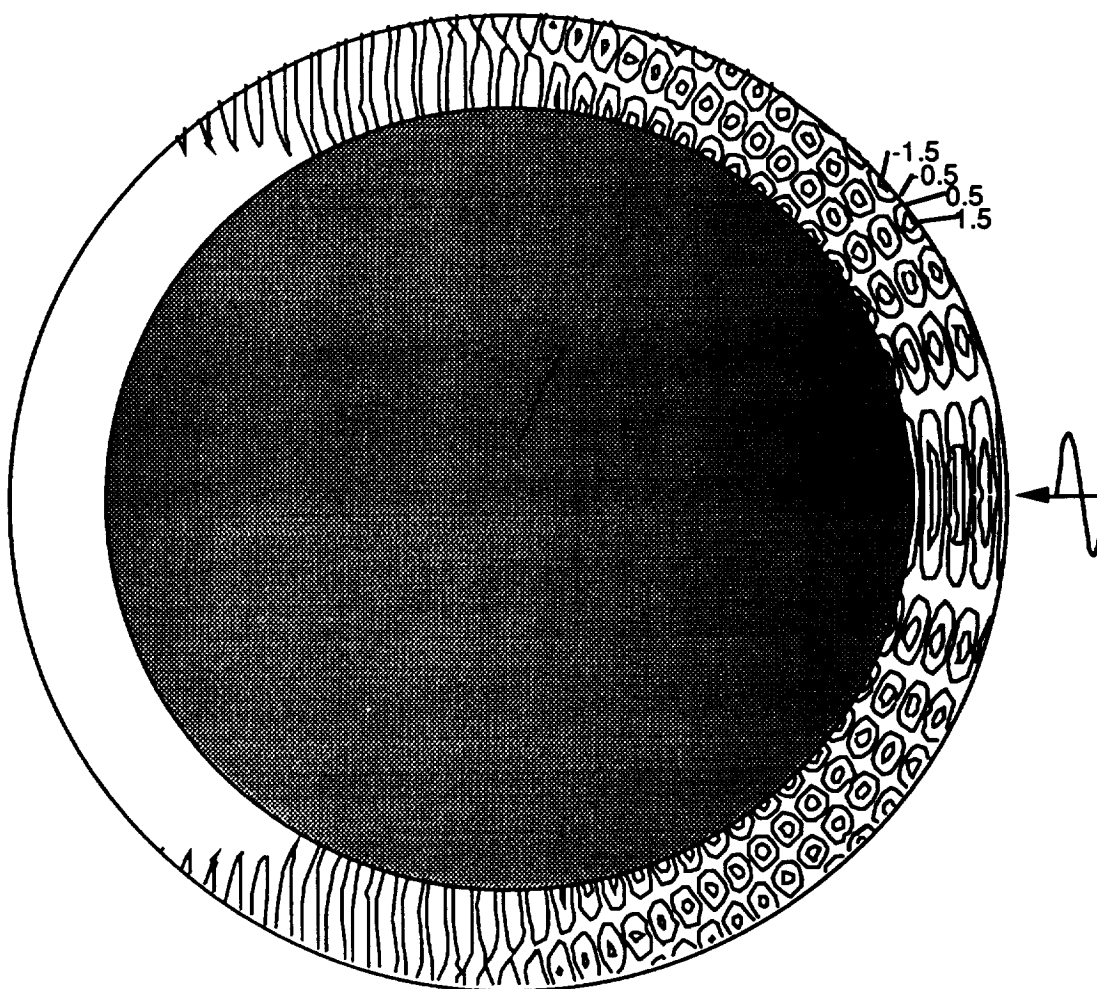


Figure 4 Contours of total field, real part for problem of scattering from  $ka = 50$  conducting cylinder, TE case. The absorbing boundary is at  $kr = 62$ , about two wavelengths from the object. The incident field wavelength is shown to scale.

equations, and the time to solve the system. An experiment with a scaled model indicates the time required for finite element models on hypercubes of differing dimensions. We construct a series of models with the number of unknowns proportional to the number of processors in hypercubes of varying dimension: 1,2,4,8,16, and 32 processors, times about 4000 unknowns. Thus the largest model has about 128,000 unknowns in this experiment. We found the time for element set-up was 16 seconds in all cases: the set-up work scales linearly with the number of unknowns, and the parallel efficiency is nearly perfect, due to the lack of communication and the balance of the load achieved by the partitioning algorithm. The time to solve the system with the iterative solver varies somewhat erratically due to the indeterminacy of the number of iterations required; however, the time per iteration was also nearly constant in all cases, about 0.5 seconds (within 0.02 seconds tolerance). Due to communications overhead, there is a slight increase with increasing hypercube dimension, so that an iteration on the 32 node hypercube took about 6% longer than an iteration on a single node (which was processing only 1/32 as many unknowns).

We are developing a full 3-D scattering code. Extending a 2-D scattering code to handle full 3-D problems is not a trivial task. All three components of either the electric field or the magnetic field must be included in the model. Absorbing boundary conditions in 3-D are substantially different from those in 2-D, and are not widely published or tested. Proper physical conditions at boundaries of conductors, dielectric and magnetic materials are more complicated. Numerical stability of the method appears to be more difficult to guarantee

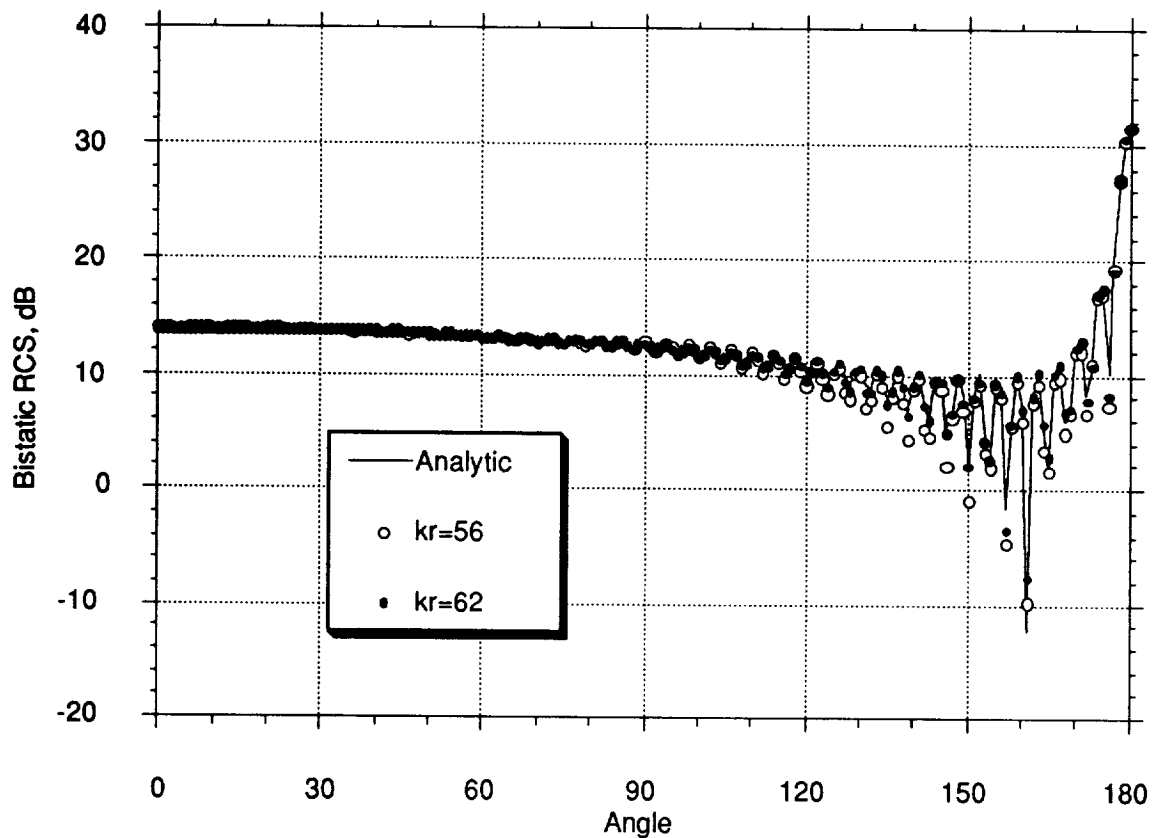


Figure 5 Bistatic RCS of TE scatter from  $ka = 50$  conducting cylinder, modeled with 4 quadratic quadrilateral elements per wavelength. Solid curve is analytic solution; open circles, finite element result for absorbing boundary radius  $kr = 56$ ; dots, finite element result for  $kr = 62$ .

than for the 2-D case, possibly requiring exotic vector basis functions and elements, such as so-called edge or tangential elements. Construction of consistent finite element meshes becomes substantially more difficult in 3-D, particularly for complicated objects containing varied materials. All of these difficulties are under investigation, and appear to be surmountable in the near future. The result will be a flexible environment for solving complicated 3-D electromagnetic scattering problems, relying heavily on the ability of parallel computers to solve such problems with hundreds of thousands of unknowns.

**A Heterogeneous Application:** The finite element code represents a large class of homogeneous problems, in which each node of the hypercube is doing the same type of task at the same time. The instructions in each node are not executed in lock-step, but essentially the same program is running in each processor, with separate data. Each processor works on the set-up at the same time, and then each processor works on its part of the system solution. We have also developed more heterogeneous parallel codes, such as a processor for synthetic aperture radar (SAR) images. In this arrangement, a 32 node hypercube is divided into 4 subcubes with 8 processors each. A separate code is loaded into each subcube, and the image sections are processed in pipeline fashion. Each subcube performs a portion of the processing functions, such as a range correlation fast Fourier transform, on part of the image, and then sends the image slice along to the next processing stage. The hypercube SAR system produces images at the rate of one 40 megabyte image every 43 seconds, utilizing the Weitek processors' ability to sustain 2.5 megaflops per second per processor.

**Conclusion:** The JPL/Caltech Mark IIIfp hypercube has proven to be an inexpensive, general purpose supercomputer. Despite the unavoidable difficulties of porting existing programs to a prototype computer,

including in this case having to deal with such issues as load balancing and communications, dozens of extremely diverse applications have been ported and are found to run efficiently on the hypercube. With expanding experience and the development of increasingly user-friendly software tools, the time required to develop hypercube applications will drop dramatically, with a corresponding explosion of applications available through public domain and commercial sources. As technology continues to provide faster microprocessors and the ability to do fast communications between processors, the potential speed of such applications on future machines patterned after the Mark IIIfp is nothing short of astonishing.

## **References:**

1. Fox, G. C., Johnson, M. A., Lyzenga, G. A., Otto, S. W., Salmon, J. K., and Walker, D. W.: Solving Problems On Concurrent Processors. Volume I. General Techniques and Regular Problems. Prentice Hall (Englewood Cliffs, New Jersey), 1988.
2. Angus, I. G., Fox, G. C., Kim, J. S., and Walker, D. W.: Solving Problems On Concurrent Processors. Volume II. Software for Concurrent Processors. Prentice Hall (Englewood Cliffs, New Jersey), 1990.
3. JPL internal document: Hypercube Project Programmer's Manual. JPL D-3220, Rev. D, 1988.
4. Calalo, R. H., Imbriale, W. A., Jacobi, N., Liewer, P. C., Lockhart, T. G., Lyzenga, G. A., Lyons, J. R., Manshadi, F., and Patterson, J. E.: Hypercube Matrix Computation Task. Report for 1986-1988. JPL Publication 88-31, 1988.

## HYPERSWITCH COMMUNICATION NETWORK

J. Peterson, M. Pniel, E. Upchurch

Center for Space Microelectronics Technology  
Jet Propulsion Laboratory,  
California Institute of Technology  
Pasadena, California 91109

**Abstract:** The Hyperswitch Communication Network (HCN) is a large-scale parallel computer prototype being developed at JPL and in collaborations with several large computer companies. These companies are planning on building commercial versions of the HCN computer. The HCN computer being designed is a message-passing multiple instruction multiple data (MIMD) computer, and offers significant advantages in price-performance ratio, reliability/availability, and manufacturing over traditional uniprocessors and bus based multiprocessors. The design of the HCN operating system is a uniquely flexible environment that combines both parallel processing and distributed processing. This programming paradigm can achieve a balance among the following competing factors: performance in processing and communications, user-friendliness, and fault tolerance. The prototype is being designed to accommodate a maximum of 64 state-of-the-art microprocessors. A full configuration will provide up to 2.6 GIP, 1.2 GFLOPS, 4 Gbytes of total node memory, 700 Mbytes/second I/O rate and 4 Gbytes/second processor-to-processor communication rate. This communication architecture extends the application of parallel systems to supercomputer problems that place heavy demand on the system for high bandwidth, low latency, and non-local communication. Hence, the classification of the HCN concept as being distributed supercomputing. This paper describes the HCN system, and reviews the performance/cost analysis and other competing factors within the system design.

The research described in this paper was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory, California Institute of Technology, and jointly sponsored by SDIO/IST and the United States Air Force, Electronics Systems Division through an agreement with the National Aeronautics and Space Administration.

**1. Applications:** The need to solve more complex problems is outpacing the ability of the world's fastest computers to solve the required applications within acceptable time periods. At the same time, even with the continuing advances in microelectronics technology, it is becoming increasingly more difficult to design and build powerful computers. The demand for increasing capability, higher performance and fault-tolerance are continually being placed on computers. Several application examples are given below.

**1.1 Space Flight Operations:** The ground based command and control operation system has played a crucial role in JPL and NASA's success. Traditional mainframe computers have been employed for science and engineering applications in the past, providing centralized processing and

data management resources for each project. The demand on computation and data handling capabilities multiplies as the complexity of spacecraft and their operation grows; the computational demands is also exacerbated when operations considers the concurrent multiple missions environment. The addition of more computers appears to be the best solution today and the present Space Flight Operations Center (SFOC) system is comprised of about one hundred conventional workstations networked together.

The following Flight Missions: Magellan, Galileo, Ulysses, Mars Observer, TOPEX, and CRAF/Cassini are expected to be operating together in the next decade; EOS, Lunar Observer, Rover, and another half dozen missions are proposed to be supported by JPL. The current Flight Operations network will eventually become "unsteerable" assuming even a factor of four increase in physical count. What is needed is a ground data system that will provide for more accurate and timely data processing that is both informative to the user and cost effective to the project.

**1.2 Consolidated Command Center:** Military applications comprises information management systems (consolidated command center) to support needs in terms of planning, decision making, and fault diagnosis. As with the JPL needs, ground data systems are needed to help cope with the increasingly complex problems in the logistics of supply and support, as well as with strategic and tactical planning. Although the precise functional and performance requirements of such a consolidated command center are of necessity evolving, certain basic, generic command center requirements include: survivability which places a geographical distribution requirement on the system; high level of fault tolerance; multi-level security; flexible and high bandwidth communications and networking; interfaces with a wide variety of machines; scalability in order to match performance requirements and dynamic reconfigurability to respond to variable workloads; high performance database management; supercomputer class floating point computation speed. Many of these requirements are also needed for the JPL Flight Operations applications. Therefore, these generic requirements are the driving force behind the HCN system described herein.

**2. SYSTEM OVERVIEW:** The HCN computer is a loosely coupled MIMD system with distributed local memories attached to multiple processor nodes, see Figure 1. The interconnect topology is a hypercube network used with a hyperswitch[1] message routing element in each node. Message passing is the major communication method among the computer nodes in the HCN computer shown in Figure 1. The HCN Message-routing method demonstrates a highly fault-tolerant capability, while providing an adaptive routing hardware based algorithm with very low message latency. With a very low communication overhead, parallelism is potentially profitable. This is because the programmer seeking maximum performance is strongly tempted to partition a problem into the finest possible granularity to create the maximum amount of parallelism. Therefore, in a message-based parallel computer, the performance of fine granularity computation depends crucially on the rate of message exchange. Fine granularity is a system which effectively supports processes transmitting short messages between code blocks that are less than several hundred instructions in length.

**2.1 Hardware:** Each node comprises one or more state-of-the-art Motorola microprocessors, and is expected to provide from 50 to 300 MIPS per node with

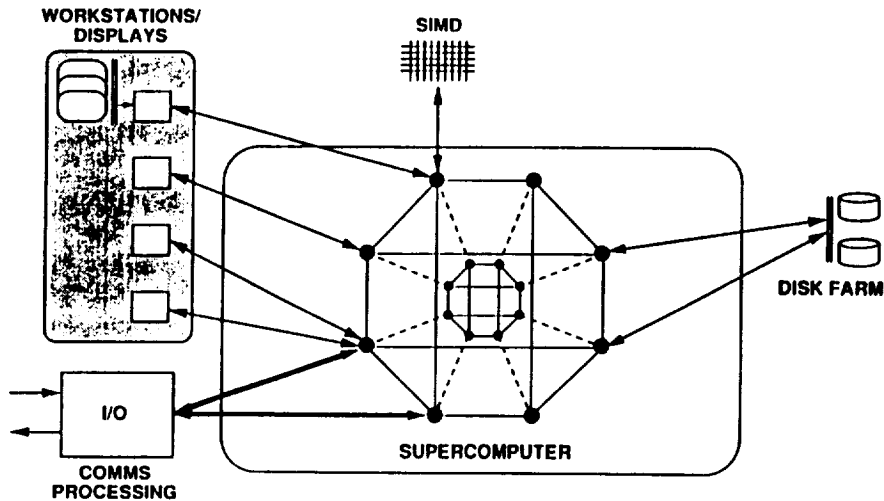


Figure 1, Hyperswitch Communication Network prototype

comparable floating-point performance. The prototype machine as a whole will comprise a total of 32-nodes, at least two microprocessors per node.

The system of links, communication, and application processors thus supplies an homogeneous, MIMD supercomputing resource, while allowing for expansibility and attachment of special-purpose processors. The partitioning of the system into communication, application, and special-purpose devices is reflected in the software: the communication system is completely hidden by the operating system software which presents a distributed object-oriented view. Applications run on the application processors and utilize the attached special-purpose processors in a transparent, fault tolerant fashion. The class/object view which hides the communication system also hides the attached devices without limiting their use.

**2.2 Software:** The class/object paradigm provides for intra-process (intra-program) communications. Global, named communications links, such as might be needed for a distributed DBMS, are easily implemented. The distributed operating system view is shown in Figure 2 as compared to a current workstation class networking view. We also expect fault-tolerance facilities to be supplied through the object system. By refining an object's simple send method to be an atomic multicast, we set the stage for shadowed processes virtually transparent to the user.

The programming environment will evolve noticeably in ten years. We see powerful debugging and monitoring tools developed which provide facilities equal to or better than those provided by the tools currently available for developing sequential programs on workstations.

**3. Architecture:** The architecture of the HCN is shown in Figure 3. This architecture is based on the new Motorola M88K microprocessor, a custom message processor and the JPL custom hyperswitch communication chips. The HCN architecture is designed to support applications that require both fault tolerance and high-performance. This architecture extends the application of



parallel computers systems to large-scale problems that place heavy demands on the communication network for high-bandwidth, low latency, and non-local communication. The overall performance of systems composed of many tightly coupled processes, such as data searching, sorting, graphics, information processing, etc. depends largely on both the efficiency of communication between nodes and the efficiency of fault recovery. All node functions support this structure.

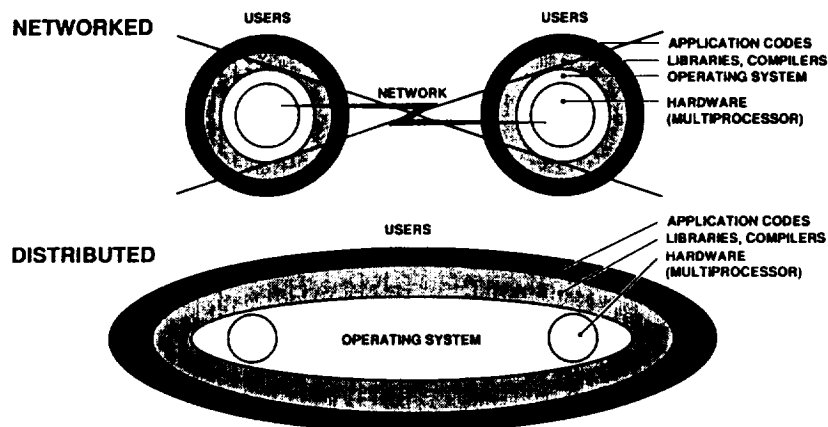


Figure 2, HCN parallel processor operating system with distributed multicomputer operating system support

**3.1 Computer Node:** The node processor structure shown in Figure 3 consists of two M88K units on one system bus (IMbus). Each unit contains one M88K microprocessor and four M88K cache and memory management devices. These devices include high-speed memory caching, two-level demand-paged memory management, and support for shared-memory multiprocessing. The M88K is a high-speed reduced instruction set computer (RISC) microprocessor. One M88K unit can be configured as the master CPU and the other M88K unit as a checker. This master/checker configuration contains comparator circuits which examine internal and external state of all active output signals. If a mismatch occurs on any output, then an error signal is asserted, the node hardware recognizes the fault and the operating system software reorganizes to eliminate a faulting node by logically enabling a redundant node. In addition to using the master/checking mode for fault-tolerant applications, the two M88K units can be used as independent microprocessors for increased performance (shared node memory multiprocessing).

Message processing latency is reduced by directly executing messages with the custom message processor using special microcode and hardwired logic. The message processor provide support for macro primitive loading and execution, message error checking, message transmission, message broadcasting, process-to-process synchronization, and message receive buffering. The message processor is also configured in a master/checker configuration for fault recognizing and recovery. Also, processor registers are used to save the message transfer control contents when other communication interrupts occur. The custom message processor is designed to field the communication interrupts and handle all the ordinary communication events.

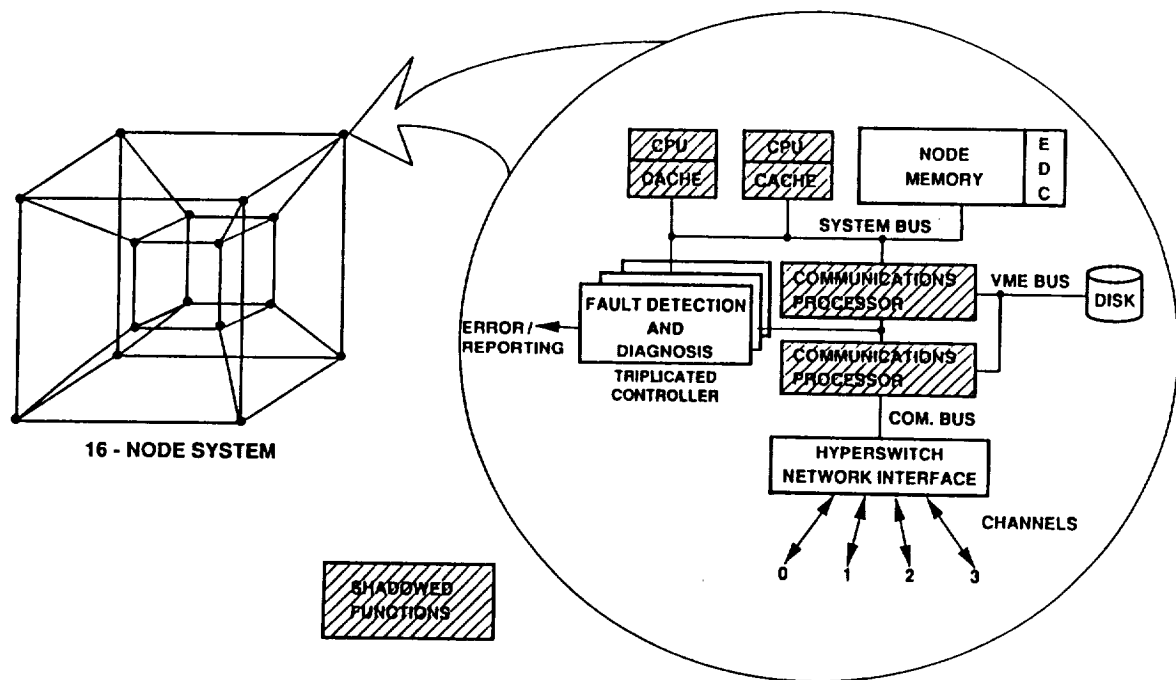


Figure 3, HCN node processor structure

In early hypercube systems message routing latency was in the hundreds of microseconds, but with the development of the hyperswitch communication chips and the message processor, latency has been reduced to a few microseconds. With a hypercube interconnection network, any two nodes that are not directly connected by a link must have their message connected by intervening nodes. But, within a hyperswitch communication network there is virtually no performance degradation when message are sent from one end of the network to the other end. The tested hyperswitch chips use an informed heuristic search algorithm, which can automatically avoid congested or faulty links based on its previous congested experience. Therefore, a message does not wait for a busy link because the hyperswitch network tries to route the message through noncongested or fault-free links. The I/O links are two 200Mbits/s bit serial channels, one for data input and one for data output. One of the  $\log n$  I/O node links can be selected as a fiber optic channels for long-haul communications. Multiple fault detection and recognition is built into the hyperswitch communication chips. This allows dynamic recovery software to reorganize around a faulting channel or node to restore normal operation.

**4. Operating System:** The HCN-based operating system (OS) will be a balance among the following competing factors: performance in processing and communications, user-friendliness, and fault tolerance. We can make best use of our resources by adopting existing operating system code wherever possible, and by building a system that supports modern programming paradigms. Figure 4 shows the user environment and concurrency support available from the OS. The primary programming paradigm to be supported is an object-oriented model, in which each object is resident on some node of the machine or distributed over several nodes for performance. Since no single programming paradigm is appropriate for all applications, a second view will be supported: processes and messages. The processes referred to here are either UNIX-style processes or lightweight tasks. While this view is closer to the view given by the JPL Mark III[2], a principal difference will be a marked

reduction in the "hypercube" view. The user will be less aware of the cube-based communications than was the case with the Mark III.

**4.1 Concurrent Object-Oriented Programming:** The HCN operating system supports an object-oriented concurrent programming paradigm. Concurrent object-oriented programming is a methodology in which the system to be constructed is modelled as a collection of concurrently executable program models called *objects*. This powerful paradigm in which the HCN is to be written exploits parallelism both in the architecture and application.

**4.2 Programming Environment:** C++ is chosen as the primary programming paradigm for the HCN because it can serve two purposes: (1) its compatibility with C makes it a language close to the machine so that all important aspects of a machine are handled simply and efficiently in a way that is reasonably obvious to the programmer. For example, the user creates an object and specifies where the object is to be placed. All subsequent manipulations of that object are done in the usual C++ fashion with no reference to the object's location. (2) the object-oriented features in C++ makes it a language close to the problem to be solved so that the concepts of a solution can be expressed directly and concisely. C++ provides constructs to express class/subclass hierarchies, type abstraction and inheritance. Extensions are added to C++ to support parallel processing, such as remote object creation and concurrent message passing using futures.

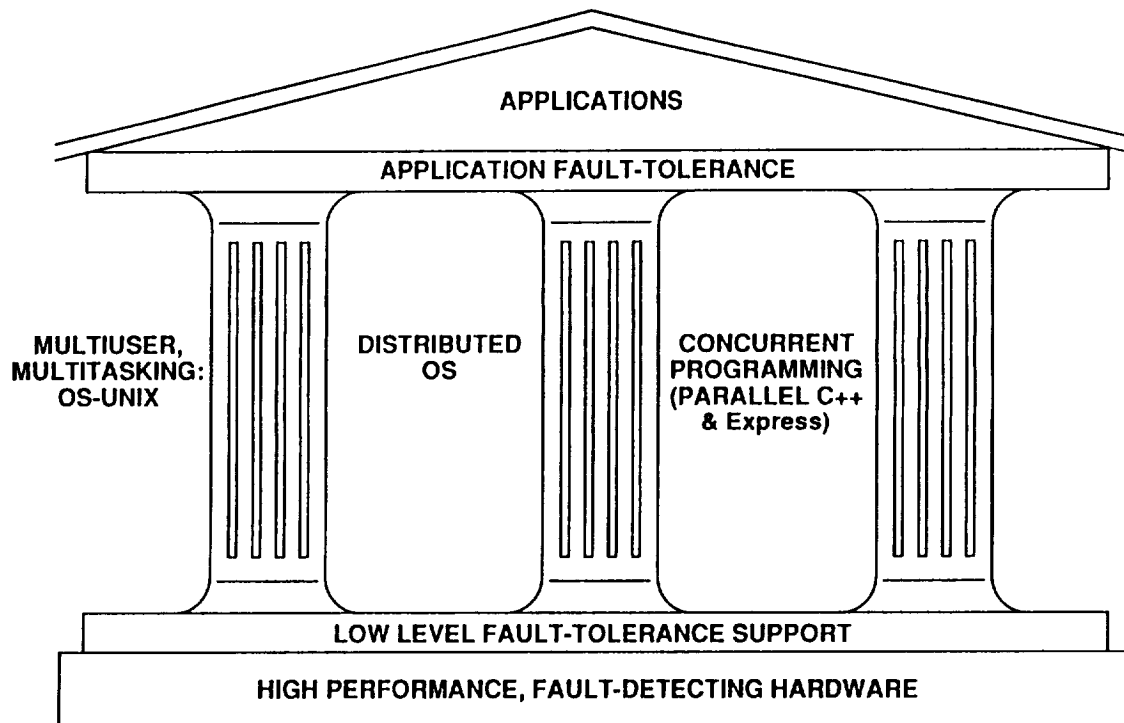


Figure 4, HCN user environment and concurrency support

**4.3 Concurrent Process-Oriented Programming:** The process-oriented programming paradigm is a more traditional way to program an application on parallel machines --- a set of sequential processes cooperatively solve a problem by exchanging information via message passing. It has shown that the process model is a powerful programming

model for a distributed-memory multiprocessor. Therefore, HCN OS will also support the process model in addition to the Object-Oriented Model. A Unix-compatible distributed operating system, such as Mach[3] or Chorus[4] is under consideration. The advantage for Unix compatibility is that much existing software can be ported to HCN easily, therefore, the application development effort will be significantly reduced. Mach is a multiprocessor operating system kernel developed at Carnegie-Mellon University. In addition to binary compatibility with Berkeley's UNIX 4.3, it also provides facilities for supporting shared-memory or distributed-memory multiprocessors, a new robust virtual memory design and a capability-based interprocess communication facility. The capability-based design and the virtual memory design in Mach can be enhanced to support mandatory access control. By supporting both Object-oriented and Process-oriented programming paradigms with Unix compatibility, fault tolerance, and multi-level security, the HCN OS is capable of serving many different types of applications.

**4.4 Process Management:** Multitasking will be fundamental to the operation of each node, as well as memory management and memory protection. This is for several reasons, including performance with an asynchronous communications system, multi-user time-sharing of the HCN, and more flexible programming for the user.

The HCN OS process management will be similar to that of Mach. The kernel will support the concept of *threads* (lightweight tasks) which allows the construction of multi-threaded tasks. Such tasks can contain multiple execution paths, all of which can be active concurrently. Threads of a single task can execute concurrently, each in a separate physical processing element. Threads may be created, terminated, suspended and resumed with HCN OS primitives that are much faster than corresponding *fork/exec* 's of UNIX.

**4.5 Memory Management:** The HCN OS will provide memory management, including virtual memory, similar to that of Mach. The kernel performs memory management at a node where physical memory is treated as a cache for the contents of virtual memory objects. In Mach each virtual memory object is managed by a *pager*. Such pagers could be used to allow memory sharing across a loosely-coupled or distributed configuration.

**4.6 Message System:** Two HCN OS message management systems will be supported. One is similar to that of ES-Kit. ES-Kit[5] which is an operating system kernel developed at MCC to support distributed, object-oriented execution in extended C++. The other is Express[6] which is a message management system that provides a portable platform on which parallel programs and applications can be built. Therefore, applications built on other concurrent computers using Express can easily port to the HCN computer. Both systems will be employed as the bases of the HCN Operating System. Efforts will be devoted in evaluating the feasibility of adding other parallel constructs, such as Distributed Objects and Multiple Threads. All communication between nodes is by messages. Therefore, both systems must provide communication services without sacrificing performance. Express offers a well understood programming model and is backwards compatible with existing applications. Whereas, ES-Kit offers a much more sophisticated development tool and a strong basis for experimental developments, such as fault-tolerant parallel extensions.

**4.7 Programming Tools:** A C++ source-level debugger aware of tasks and remote objects, and performance-monitoring tools for visualizing program behavior will be available. The emphasis will be on graphic tools and a simulation environment for the debugging of application code. Graphical tools such as a hierarchical diagram of classes and instances, for example, greatly increase program reliability and programmer productivity.

**4.8 Distributed File System:** HCN OS supports transparent remote file access whether the file resides on a disk attached to a remote HCN node or on a Workstation connected to the HCN network. The file system and directory structure is Unix compatible. For example, facilities are provided to mount/dismount file systems, and for file transfers between different disk drives.

**5. Fault-Tolerance:** The HCN is a set of homogeneous processing elements interconnected by a high bandwidth network. These processing elements are connected to a heterogeneous set of data sources and sinks, including: workstations; graphic displays; disks; and special purpose processors. In order to make the entire HCN a fault-tolerant system, we need to assure fault-tolerant operation in the following three components:

1. The homogeneous processing element has to be fault-tolerant.
2. The communication network has to be fault-tolerant.
3. The interface to the heterogeneous external sources or sinks must be fault-tolerant.

The fault tolerance design of HCN should be able to survive one or more failures falling into any of the above three categories. In the prototype effort, HCN is used as the homogeneous processor network. In the following paragraphs, we will discuss the fault-tolerance design in the HCN with respect to the above.

1. Each HCN node has built-in self-checking hardware consisting of dual Motorola 88000 CPUs for error detection, and error detection circuitry for node memory and system buses. An exception is signalled to the operating system when an inconsistency is detected by the hardware. The OS then initializes the damage assessment program and error recovery program to identify the type and location of the fault, and resumes the whole system to a safe state. In a distributed system, a single node failure cannot be isolated from the rest of the system. Therefore, a global recovery mechanism has to be employed to synchronize and reconfigure the system. The common checkpoint/rollback recovery technique has been identified as inadequate for a message-passing distributed system due to the so-called "domino effect". Instead, user processes are duplicated in two different nodes and processes are synchronized actively by messaging calls. When the primary process is faulty, the backup process will resume the primary's position with minimum recovery delay. This fault tolerance capability is transparent to the user.
2. The Hyperswitch can detect channel errors by two levels of parity check. The adaptive routing algorithm built into hardware can then bypass faulty links and route a message through. It can also recover transient errors occurring in data transmission by automatic retry. Also, self-timing

hyperswitch channels can determine the data rate locally, no system-wide clock is needed. However, the Hyperswitch has limited hardware support for message broadcasting, not to mention atomic broadcasting. Moreover, it does not search exhaustively all the possible routes, and thus may not be able to find a route successfully in the presence of faulty links. The HCN operating system has to perform the following functions to augment the fault tolerance abilities of the hardware:

- To be able to send point-to-point messages in spite of faulty links or nodes in the hyperswitch network.
  - To be able to tolerate up to  $(n/2)$  link faults, where  $n$  is the dimension of the cube. In other words, a point-to-point message can be routed between any two pairs of nodes if there exists less than or equal to  $(n/2)$  faulty links.
  - Should be able to find a feasible minimal path when the hyperswitch fails to find an optimal route when faulty links exist.
  - To be able to broadcast message to the entire cube or to an arbitrary set of nodes
  - The broadcast message must either be received by all the non-faulty nodes in the recipient group or none of them (i.e., atomic broadcasting)
  - To be able to use a fault-tolerant broadcasting algorithm to bypass faulty nodes and links in building minimal spanning trees.
3. When an HCN node is connected to an external device via a channel, say, a VME interface, either the node, the channel, or the external device may cause single-point failure to the entire system. Therefore, all the three components have to be protected from failure with redundancy. In addition, a fault-tolerant interface between the HCN node and the attached devices/processes has to be built to perform error recovery for the external device. For a device without self error-checking capability, triple module redundancy may be adopted and a voting mechanism has to be included in the interface software/hardware. For a device with self checking capability, a duplicate device is necessary for backup purpose. The physical channels between the node and the external device should also be duplicated to protect the system from single-point failure on the node that is attached to the external device.

**6. Cost/Performance Tradeoffs:** With today's microelectronic devices the cost of fast devices tends to grow faster than the performance benefit of the increased device speed. Hence, the cost per unit of computing power tends to be greater for high-end machines than for low-end machines, although this trend is technology dependent and could change over time. The relative performances and cost ranges of four classes of commercial computers[7,8] are plotted in Figure 5. The estimated performance and cost ranges of the HCN is also shown. As you can see the low-cost technology of the HCN is an opportunity to create a cost-effective high-performance system by combining slow-speed microprocessors. As stated in Section 2., the cost advantage of using low-cost technology is balanced by the degradation in efficiency that inevitably occurs as the number of processors increases.

Therefore, Communication efficiency and hardware connectivity are the major concerns in the choice of a cost-effective message-passing computer architecture.

To address these concerns, system modeling is used to scientifically explore cost/performance tradeoffs of the HCN-based system[9]. The basic modeling approach is part of a design methodology that has been called performance engineering and uses hierarchical modeling with models expressed by extended directed graphs. High level models are used to evaluate design tradeoffs which includes applications software and operating systems as well as communications hardware overheads. The model can be driven either by synthetic scaled workloads or applications traces.

Some of the design questions being explored with the model include: optimal mapping of application functions to hypercube nodes with and without system faults; sensitivity to message/packet length and network topology; operating systems overhead for various message protocols and data checkpoints; effects of adaptive routing with high volume traffic and bursty traffic.

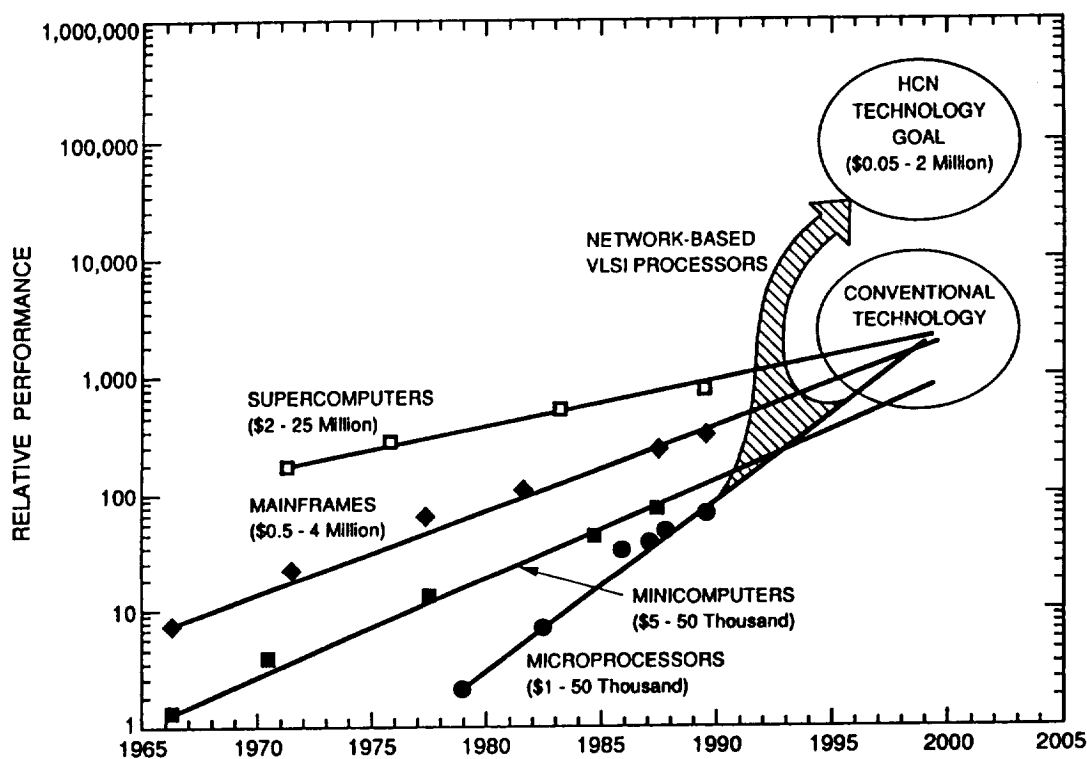


Figure 5, Computer systems cost/performance as a function of trends

**7. Conclusion:** An important part of this work included establishing a strategy for how long-lived systems should be designed and constructed. In particular, one should view the HCN as an ongoing and continuing design -- the HCN is never complete in the sense that newer and better technology continues to appear, and it must be able to take advantage of that technology. Furthermore, the users of the HCN systems are also changing and improving its utilization -- that is, new applications are encountered and new responses to those applications must be devised and implemented. Therefore, as work

gets underway on the HCN, it is important to be designing the next step. Of course, the next step is not a complete replacement, but is an evolution of its fundamental components. By the time several steps of that evolution have occurred, the system may be quite different from its original form, but better and more adapted to the problems that it solves.

Given the rapid pace of technology, a design that reaches ten years into the future can serve best by being flexible -- able to incorporate modification, extension, and, basically, to evolve. In such flexibility there is strength.

## References:

1. Chow, E., Peterson, J., Grunwald, D., and Reed, D., "Hyperswitch Network For The Hypercube Computer", Proc. 15th Conf. On Computer Architecture, May 1988.
2. Peterson, J., Tuazon, J., Pniel, M., Lieberman, D., "The MarkIII Hypercube Ensemble Concurrent Computer", Proc. of the 1985 International Conf. on Parallel Processing.
3. Acceta, M., and et. al. , "Mach: A New Kernel Foundation For UNIX Development", Computer Science Department, Carnegie-Mellon University, May 1986.
4. Rozier, M, et. al., "CHORUS Distributed Operating System", Technical Report CS/TR-88-7.6, Chorus Systemes, November 1988.
5. Leddy, W. and Smith, S., "The Design of the Experimental System Kernel" MCC TR# ACA-ESP-089-89, March 1989.
6. ParaSoft Inc., Express Users Manual, Pasadena, Calif., 1990.
7. Hwang and Degroot, Parallel Processing for Supercomputers & Artificial Intelligence, McGraw-Hill, 1989.
8. Patterson and Hennessy, Computer Architecture - A Quantitive Approach, Morgan Kaufmann, 1990.
9. Upchurch, E. and Neuse, D., "Modeling Hypercube Communications Network", Proc. Summer Simulation Conf., Austin, July 1989.



Biological Neural Networks As Model Systems For Designing  
Future Parallel Processing Computers

Muriel D. Ross  
NASA-Ames Research Center  
Moffett Field, CA 94035

One of the more interesting debates of the present day centers on whether human intelligence can be simulated by computer. Some [1-3] argue that there is something unique, or even mystical, about the human mind, so that intelligent behavior cannot be simulated by machine [1, 2]. Others [4] disagree on the premise that brains, on which minds depend, are simply collections of neurons and, since neurons are physical entities, they can be understood in physical and even mathematical terms, perhaps not soon, but in the foreseeable future. In an interesting discussion of this debate, Denning [5] points out that regardless of the questions raised by philosophers, investigators will continue to explore the possibility of producing thinking machines by building increasingly sophisticated systems.

In my laboratory, we work under the premise that neurons individually are not smart at all. Rather, they are physical units which are impinged upon continuously by other matter that influences the direction of voltage shifts across the units' membranes. Some influences are in the direction of depolarization (excitation) and others are in the direction of hyperpolarization (inhibition). In one type of neuron, the machinery of the cell responds in analog fashion to these incoming signals as spatio-temporal events that may or may not result in overcoming a threshold. If threshold is overcome, the cell responds digitally by initiating patterned, electrical impulses that are conducted along the axon to its terminals, usually some distance away. There, the impulses result in probabilistic, quantal release of neurotransmitter [6] at one or more synaptic sites on other neurons. The specific actions of the various spatio-temporal inputs on the postsynaptic cells again determine their responses. In a different type of neuron, the cell does not discharge impulses but spreads the response electrotonically along its membrane to synaptic sites. Neurons of this type function locally in the network and appear to be extremely important in information processing.

It is only in the actions of a great many neurons, billions in the case of the human nervous system, that intelligent behavior emerges. The enormous number of constituent parts and the complexity of the connectivities are the bases for doubts that the nervous system can be understood and simulated by computer. A common argument is that we would gain more insights by approximating biological neural networks as we think they might exist rather than by attempting to learn from them by rigorous study.

The problem with this approach is that, in the end, it will not bring us to our ultimate goal. Extrapolating from generalization to generalization removes us that much farther from understanding the bases for biological system robustness, memory, learning and intelligent behavior. Such extrapolations may produce more powerful machines than exist today, as Denning [5] has suggested, but they will fall far short of the desired end point of a truly thinking machine.

What is required to understand even the simplest neural system is a painstaking analysis, bit by bit, of the architecture and the physiological functioning of its various parts. The goals should be to gain insight to those features that are fundamental to all neural networks and then to develop the algorithms that define the responses. This kind of effort will result in new applications of biological system attributes to artificial systems and will result in advances in computer design. The research does, however, require the interaction of neurobiologists, neurophysiologists, physicists, modelers, mathematicians and theoreticians; in short, a formidable array of specialists. It also requires the development of highly specialized computer technologies. At the NASA-Ames Biocomputation Center, we have a team assembled to promote the technology required to understand the neurobiology, both in its own right and as a guide to future computer advances. We are also working to achieve the physical and mathematical interpretations essential to develop a silicon chip for potential inclusion in robots.

The biological neural networks we study, the vestibular utricular and saccular maculas of the inner ear, are among the most simple of the mammalian neural networks to understand and model. At the same time, they are complex enough to be useful for deriving principles of neural network organization. Because the basic organization of vestibular maculas does not differ between species, we have used the rat utricular macula as the model mammalian system.

Macular endorgans function as linear bioaccelerometers. Not surprising to engineers, maculas consist of a test mass above a detecting unit. The test mass, suspended in a weak gel-like liquid, consists of tiny crystallite particles (otoconia) that are unevenly distributed above the detector. The underlying detecting unit is a neural network structurally organized for weighted, parallel distributed processing of information. The network consists of two kinds of hair cells (type I and type II), a system of nerve endings with branches and collaterals, and a number of small-diameter nerve fibers that end on type II hair cells and on other neural elements within the macula. The network is not identical in detailed organization from site to site, and there may be a match between otoconial loading and network properties [7]. If so, parallel processing of acceleratory information begins at the input, otoconial layer even though it is a non-neural entity.

The hair cells function as detectors. They bear a tuft of thread-like processes, called stereocilia, and a single kinocilium at their apical surfaces. The stereociliary tufts are always organized in hexagonal arrays, but the size and height of the individual stereocilia, as well as their number, differ from site to site. Ordinarily, the stereocilia are in staircased order, with the tallest bordering the kinocilium, which is the tallest of all. The kinocilium is attached to the otoconial layer by strands of organic substance. Since the kinocilium has been shown to be capable of motility [8], we believe it agitates the otoconial layer. Translational linear acceleratory force affects the otoconia differently, according to their individual masses and the background of activity already going on. The result is that complex waves are constantly being emitted, with or without the addition of transient accelerations, and the detecting units respond according to their stereociliary configurations.

Fourier analysis of various tuft organizations shows that the stereociliary tufts are highly directionally tuned by their hexagonal organization and their specific stereociliary heights [9]. In another astonishing correlation to human-engineered devices, it seems that nature invented the equivalent of a sensitive phased array antenna millions of years ago! Analysis of the repeating lattice angles of stereociliary tufts demonstrate that they range between  $115^{\circ}$  and  $125^{\circ}$  [10, 11]. The optimum repeating lattice angle for human-engineered antennas is a  $120^{\circ}$  rhombus [12].

These and other findings mean that we are well on our way toward understanding, and expressing in mathematical and engineering terms, the basis of detection of incoming linear acceleratory signals by this biological system. Achieving an understanding of the neural network organization is also well underway, using computer technology and specially developed software [13] to reconstruct actual parts of the network and to produce symbolic models to mimic their functioning. This research currently is tedious and labor-intensive, requiring photography of serial sections in a transmission electron microscope, reassembly of the micrographs into montages of the sections, tracing objects of interest from the montages, digitizing the tracings into a computer, and reassembling the tracings into shaded solid and transparent images. The research would be greatly expedited by semiautomation of data collection to eliminate photography completely. At NASA's Biocomputation Center, we are working to advance this technology. Achieving the goal of semiautomatic data collection and reassembly will be a breakthrough for neuroscientists trying to understand the architectures of more advanced biological neural networks and for scientists wishing to apply insights obtained from biological systems to advance parallel processing computers and robotic devices.

Using current reconstruction techniques, we have been able to demonstrate the smallest functional units of the network, its receptive fields, which consists of the calyceal ending(s) of a nerve fiber together with all the hair cells that synapse with it (them). We have also reconstructed small parts of the neural network. The basic findings are that no two of the receptive fields are identical and that the network varies in complexity from site to site [14, 15].

Type I cells are enclosed by the calyces and synapse only with them. Type II cells lie outside the calyces and distribute their output to as many as four neighboring calyces by synapsing either with the calyces or with calyceal collateral processes. Calyceal collaterals are either presynaptic (feedforward), postsynaptic (feedback), or reciprocally (bidirectional) synaptic to type II cells. Feedforward collaterals end opposite subsynaptic cisterns, strongly suggesting that the collaterals have an inhibitory action (they hyperpolarize the type II hair cell) (discussed in [16]). Experimental evidence [17, 18] indicates that synaptic junctions with subsynaptic cisterns function in long-lasting hyperpolarization of the postsynaptic cell. The findings strongly suggest that type II hair cell activity is dampened when calyces and their processes are activated

(depolarized), possibly providing for a center-on surround-off response that is well known in the retinal neural network.

In addition to the vestibular nerve fibers just described, there is a system of small, beaded fibers (efferents) that arises extrinsic to the macula and is presynaptic to all the other neural elements except for type I hair cells. The nerve fibers wander through the macular neuroepithelium where their beads, more accurately their "boutons", synapse with type II hair cells, calyces, collaterals, nerve branches and branch junctions. Morphological and experimental evidence strongly suggests that the endings on type II hair cells, which have subsynaptic cisterns, provide a background of inhibition (hyperpolarization) of the type II hair cell, while the remaining endings on other neural elements provide background excitation (see [16] and [17]). One bouton may synapse with both a type II hair cell and a calyx, so that the specific function of the synapse appears to be determined by conditions on the postsynaptic side.

In a broader context, because all the synapses likely release transmitter substances quantally and perhaps with nonuniform release probabilities [6], the dynamics of the functioning network are likely nonlinear. The salient question is whether we can reduce the complexities of the neural network just briefly described to meaningful engineering and mathematical expressions that capture, uncompromisingly, the architectural and functional foundations of the system without simply reproducing it. For if we are to improve artificial computer technologies and produce thinking machines, we must be able to accomplish this without replicating the billions of neurons and their connectivities present in a human brain. We must reduce brain organization to its basic organization and build from there. Perhaps in some ways the artificial brains can be made to outstrip the biological!

As part of this effort, we have produced a dynamic, symbolic model [18] of the functioning of a small part of the neural network (Figure 1). This model is based on mathematical interpretations of the neural geometry revealed by the reconstruction work. In the initial, six-tiered model shown here, three different geometric arrangements were incorporated in a two-dimensional model. The layers depicted in order are 1) stimulus, 2) type I hair cells, 3) type II hair cells, 4) calyces, 5) impulse initiation zone, 6) neuronal discharge. The output, patterned neural discharge is also displayed on the screen. On the computer, dynamic states are depicted in a range of colors, from blue (maximal inhibition) through green (neutral) to red (maximal excitation). Comparable states in Figure 1 are shown as shaded symbols (see legend). The effects of varying individual parameters, such as direction of input, hair cell polarization, length of nerve branches, and resting discharge rate, can be determined quantitatively and qualitatively by the model. Feedforward and feedback loops are now being added to study the effect of lateral inhibition on the output of the various receptive fields. Still another addition is conversion to a three-dimensional model and putting in place the background activity imposed by the small-diameter efferents.

While much remains to be learned from more advanced versions of the model, we already have found that inhibition is important in a neural network. When all elements are excitatory, the simulated network quickly saturates so that all units stay on continuously. Only when inhibition is introduced, in our model at the level of the type II hair cell, does neural coding appear. It is also apparent from the model that branching over short distances makes the unit more sensitive to different inputs at its calyceal sites (unit b in Figure 1), leading to a more irregular discharge pattern. Lengthening the branches tends to make the unit more regular in its responses.

In complementary models, we have been able to reproduce the receptive fields of vestibular maculas by the Monte Carlo simulation method [16]. That is, we derived constraints such as the number of type I and type II hair cells and the number and lengths of the nerve fiber branches in the various kinds of fields. Probability tables derived from these constraints were used together with a random number generator to reproduce examples of biological receptive fields by computer. The results of this method strongly suggest to us that biological neural networks develop by constrained randomness in wiring and are not entirely genetically determined. In future research, we hope to design artificial neural networks entirely by constrained randomness to learn whether a degree of randomness in wiring might be beneficial to artificial systems.

As we continue to develop our simulations, we also are beginning to reduce the macular architecture to symbolic block diagrams. Figure 2 illustrates one of these diagrams, for a highly branched nerve fiber. Next, these block diagrams will be reduced to electronic circuit diagrams, using small and medium scale

integrated circuit devices (SSI and MSI), for implementation in hardware. Development of a prototype linear accelerometer chip based on its biological counterpart should be possible in the next few years.

In the meantime, what are the particular attributes of biological neural networks that could help us to design better parallel processing computers? One major observation is that the nervous system appears to be organized on the basis of two circuits: highly channeled and distributed modifying [16] (Figure 3). That is, there are highly directed [19] inputs to an area, carrying highly specific information, but the information is processed further by local or intrinsic circuits that modify the output to another station. The first to suggest such a basic organization was Shepard [20], who used the less definitive terms "vertical" and "horizontal" for the circuits. Because we now find that part of the distributed circuit functions to dampen type II hair cell activity following calyceal voltage changes, it is possible that lateral inhibition is also a fundamental feature of distributed modifying circuits (see also [18]).

Our research suggests, however, that a third kind of circuit may exist. This is a system of nerve fibers that comes into an area or layer from another site and provides a widely distributed, biased background of activity against which the channeled and distributed modifying circuits operate. In natural systems, it is possible to raise or lower background activity hormonally, by increased input to it or through local, reciprocal control as occurs to some extent in the gravity receptor. That is, heightened activity in a neuron could release neurotransmitter to terminals of reciprocal type, activating them to modulate the responses of more distant neurons. This kind of circuit might prove essential to learning, cognition and creativity, as it could help focus attention to certain incoming information, correlate neural activity over variable distances, and result in disregard of other simultaneous inputs that might otherwise be distracting. In the human brain, such circuits may provide the emotional context within which higher cognitive functioning takes place (see also discussion in [3]). While for now the conservative approach is to consider this group of fibers as part of the distributed modifying circuitry, the background activity these fibers provide may prove sufficiently significant to demand recognition as a separate, third kind of fundamental circuitry.

Other factors that appear to be important for incorporation into artificial neural networks and computer technologies are the following: 1) Biological neural networks are not modular. Variation in receptive field and network organization is apparently important to parallel processing of incoming information, to segregate discrete parts of the message for different distributions. 2) Biological systems are robust and tolerant of neuron failure. Robustness is achieved in part through redundancy, but likely also through non-modularity and constrained randomness in wiring. 3) Individual neurons are tiny bits in the machine of intelligence. They are not smart enough to be able to tell another neuron that it is responding in error. Thus, backpropagation as described and utilized in engineered neural networks does not exist in biological systems. However, feedforward-feedback loops and lateral inhibition may be essential to neural functioning. 4) Constrained randomness in wiring may be typical of all advanced biological neural networks. This and other probabilistic features of biological neural information processing systems suggest that natural neural networks are dynamically nonlinear.

While we still have a long way to go to understand even this most simple mammalian neural network in sufficient detail for extrapolation to computers and robots, a start has been made. Moreover, the insights we are obtaining and the technologies we are developing should help advance our understanding of the more complex neural networks that underlie human intelligence. Progress in this area should, as Denning [5] suggests, lead to ever more sophisticated machines.

## REFERENCES

- 1) Penrose, R.: The Emperor's New Mind. Oxford Univ. Press (Oxford), 1989.
- 2) Searle, J.R.: Is the Brain's Mind a Computer Program? Sci. Am. vol. 262, no.1, 1990, pp. 26-31.
- 3) Popper, K ; and Eccles, J.: The Self and Its Brain. Springer Internat. (Berlin), 1985.
- 4) Churchland, P.M. and Churchland, P.S.: Could a Machine Think? Sci. Am. vol. 262, No.1, pp. 32-37.

- 5) Denning, P.J.: The Science of Computing: Is Thinking Computable? *Am. Sci.* vol. 78, pp. 100-102.
- 6) Walmsley B.; Edwards, F.R.; and Tracey, D.J.: Nonuniform Release Probabilities Underlie Quantal Synaptic Transmission at a Mammalian Excitatory Central Synapse. *J. Neurophys.* vol. 60, no. 3, 1988, pp. 889-908.
- 7) Lorente de No, R.: Ausgewählte Kapitel aus der vergleichenden Physiologie des Labyrinthes. *Ergebn Physiol.* vol. 32, 1931, pp 73-242.
- 8) Flock, A.; Flock, B.; and Murray, E.: Studies on the Sensory Hairs of Receptor Cells in the Inner Ear. *Acta Otolaryngol. (Stockh.)* vol. 83, 1977, pp. 85-91.
- 9) Mugler, D.; and Ross, M.D.: Phased Array Characteristics and the Directional Sensitivity of Vestibular Hair Cells. *Proc. IEEE/EMBS, Philadelphia, 1990*, In press.
- 10) Ross, M.D.; Meyer, G.; Cutler, L.; Lam, T.; and Mugler, D.: Computer Visualization Techniques Applied to Vestibular Research, *Proc. IEEE*, vol. 11, 1989, pp. 1922-3.
- 11) Mugler, D.; and Ross, M.D.: Vestibular Receptor Cells and Signal Detection: Bioaccelerometers and the Hexagonal Sampling of 2-D Signals. *Math. Comput. Modell.* vol. 13, no. 2, 1990, pp. 85-92.
- 12) Petersen, D.P.; and Middleton, D.: Sampling and reconstruction of Wave-Number-Limited Functions in N-Dimensional Euclidian Spaces. *Information and Control*, vol. 5, 1962, pp. 279-323.
- 13) Ross, M.D.; Cheng, R.; and Lam, T.: ROSS 2.3 version of CARMA, Computer-Aided Reconstruction of Macular Accelerometers. 1989.
- 14) Ross, M.D.; Cutler, L.; Meyer, G.; Lam, T.; and Vaziri, P.: 3-D Components of a Biological Neural Network Visualized in Computer Generated Imagery. I. Receptive Field Organization. *Acta Otolaryngol. (Stockh.)* vol 109, 1990, 83-92.
- 15) Ross, M.D.; Meyer, G.; Lam, T.; Cutler, L.; and Vaziri, P.: 3-D Components of a Biological Neural Network Visualized in Computer Generated Imagery. II. Macular Neural Network Organization. *Acta Otolaryngol. (Stockh.)* 1990, pp. 235-244.
- 16) Ross, M.D.; Cutler, L.; Doshay, D.; Cheng, R.; and Naddaf, A.: A New Theory of Macular Organization Based on Computer-Assisted 3-D Reconstruction, Monte Carlo Simulation and Symbolic Modeling of Vestibular Maculas. *Proc. Barany Soc.*, In press.
- 17) Shigemoto, T.; and Ohmori, H.J.: Muscarinic Agonists and ATP Increase the Intracellular  $Ca^{++}$  Concentration in Chick Cochlear Hair Cells. *Physiol. (Lond)* vol 420, 1990, 127-148.
- 18) Ross, M.D.; Dayhoff, J.; and Mugler, D.H.: Toward Modeling a Dynamic Neural Network. *Math. Comput. Modell.* vol. 13, no. 7, 1990, pp. 97-106.
- 19) Schmitt, F.O.: The Role of Structural, Electrical and Chemical Circuitry in Brain Function. In: Schmitt, F.O.; and Worden, F.G. (eds), *The Neurosciences: Fourth Study Program*. MIT Press (Cambridge, MA), 1979, pp. 5-20.
- 20) Shepard, G.M.: The Olfactory Bulb as a Simple Cortical System: Experimental Analysis and Functional Implications. In: Schmitt, F.O. (ed.), *The Neurosciences: Second Study Program*. Rockefeller Press (New York) 1970, pp. 539-552.

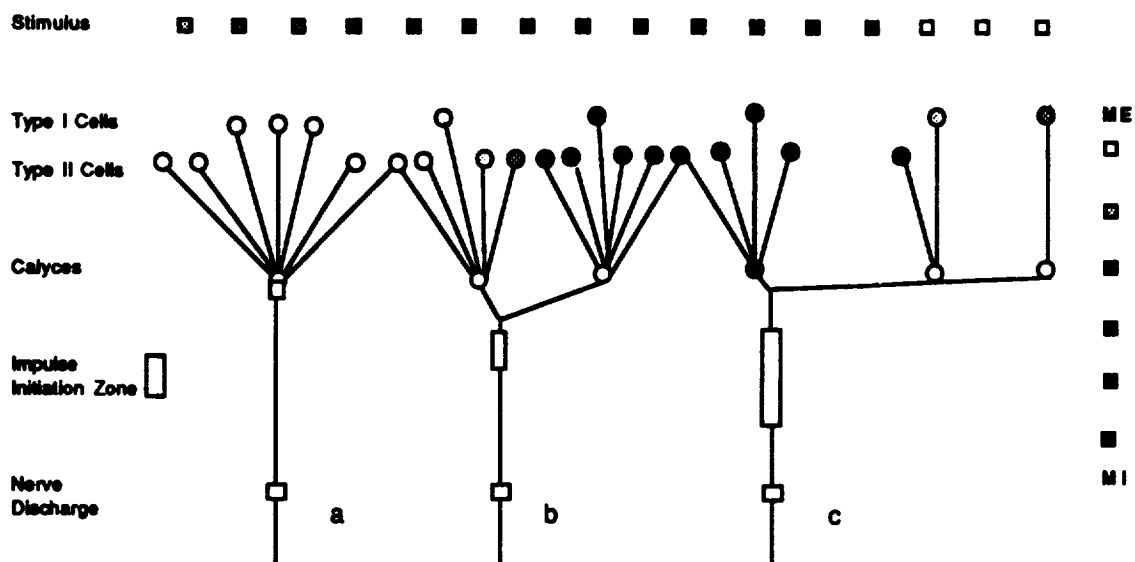


Figure 1.

This figure illustrates our dynamic, symbolic model of a functioning macular neural network. See text and legends at right and left for explanation of the model. MI, Maximum inhibition; ME, Maximum excitation.

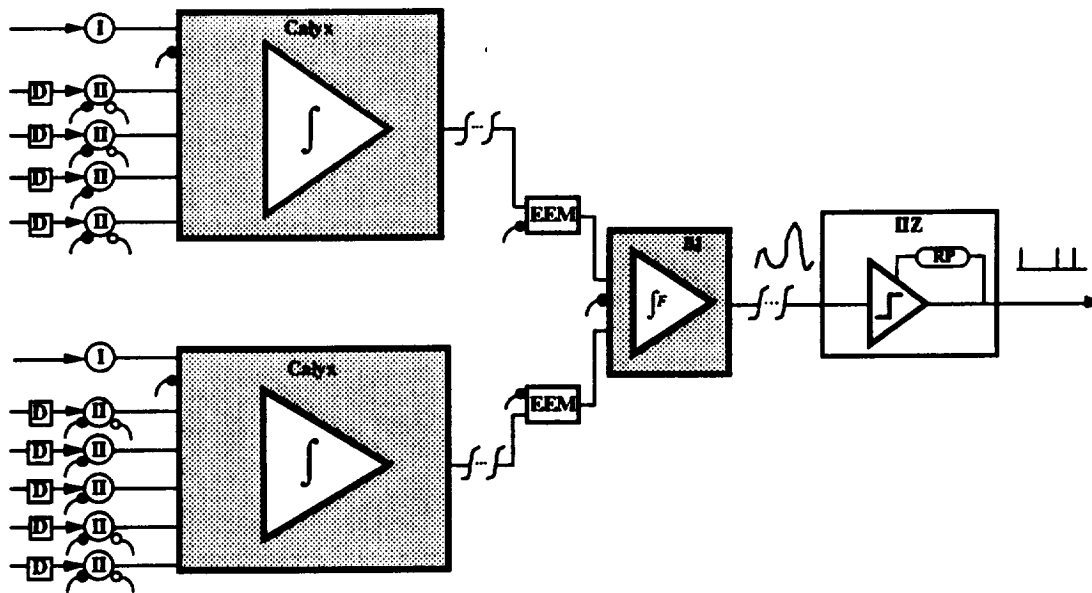


Figure 2.

Figure 2 is a symbolic block diagram of a macular nerve fiber with two branches. The symbols are: I, type I hair cell; II, type II hair cell; filled circles at the ends of curved lines, efferent terminals of extrinsic origin; open circles at the ends of curved lines, intrinsic efferent endings; BJ, branch junction; D, delay; EEM, extrinsic efferent modification of voltage spreading along the nerve fiber; IIZ, impulse initiation zone; RP, refractory period; /.../ more of the same. The black lines represent the nerve fiber and its branches (like wires, except for the slow propagation time). The integrator symbol inside calyceal triangles indicates simple integration of the inputs; an integrator symbol plus  $F$  (in BJ) indicates integration of a function of the inputs. The symbol inside the triangle at IIZ represents a threshold comparator. The undulating line above the nerve fiber between BJ and IIZ indicates that variable continuous voltage, an analog signal, is carried by the nerve fiber up to this point. The analog signals are converted to pulse trains at the IIZ. These are digital signals (pulse trains at far right).

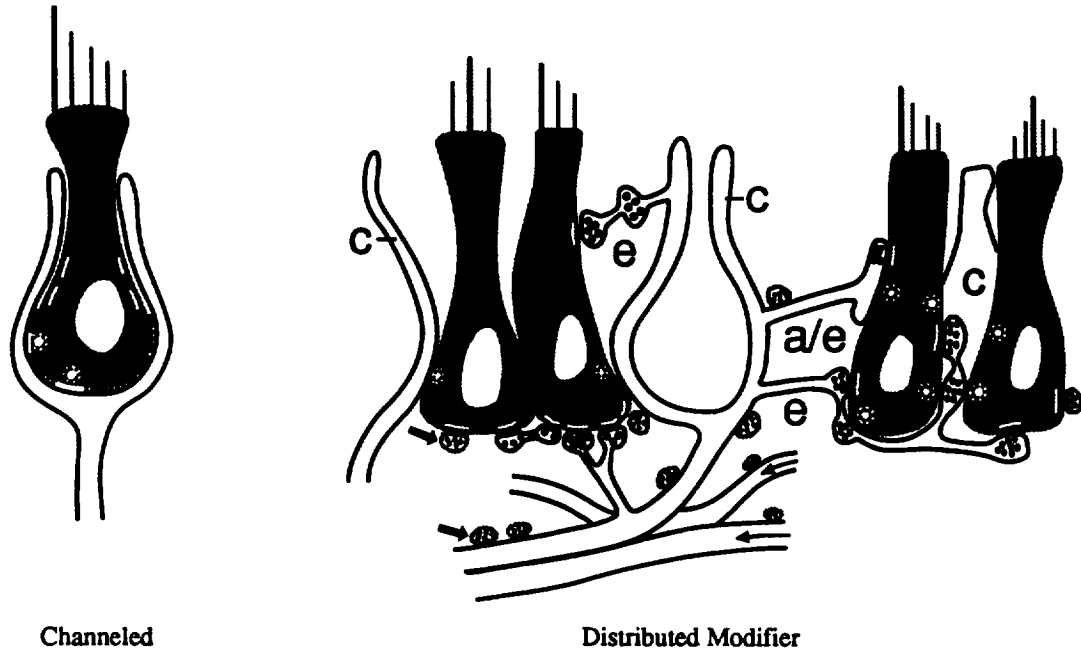


Figure 3.

This diagram illustrates the concept that certain inputs to a neural network are highly channeled, or directed, to a second unit while other parts of the network are concerned with distributing modulatory effects to nearby and more distant units. In this way, the output of the system is modified. I, type I hair cell; II, type II hair cell; a/e, afferent/efferent, or reciprocal endings; c, calyx; e, efferent-type collaterals; asterisk-like symbols, ribbon and spherular synaptic junctions; clear bars in type II and type I hair cells, subsynaptic and subsurface cisterns, respectively. Fine arrows show direction of information flow; heavy arrows indicate efferent terminals of extrinsic origin.



## **SESSION B - HUMAN FACTORS ENGINEERING AND LIFE SCIENCES (PART 1)**

**Tuesday November 27, 1990**

- **Biomedical Applications Of NASA Technology**
- **Direction-Discriminating Hearing Aid System**
- **X-Ray Imaging Microscope For Cancer Research**
- **Mechanical Response Tissue Analyzer For Estimating Bone Strength**
- **Adaptation Of NASA Technology For The Optimization Of Orthopedic Knee Implants**



## BIOMEDICAL APPLICATIONS OF NASA TECHNOLOGY

Donald S. Friedman  
Chief, Office of Commercial Programs  
NASA/Goddard Space Flight Center  
Greenbelt, MD

## ABSTRACT

Through the active transfer of technology, the National Aeronautics and Space Administration (NASA) Technology Utilization (TU) Program assists private companies, associations, and government agencies to make effective use of NASA's technological resources to improve U.S. economic competitiveness and to provide societal benefit. Aerospace technology from areas such as digital image processing, space medicine and biology, microelectronics, optics and electro-optics, and ultrasonic imaging have found many secondary applications in medicine. Examples of technology spinoffs are briefly discussed to illustrate the benefits realized through adaptation of aerospace technology to solve health care problems. Successful implementation of new technologies increasingly requires the collaboration of industry, universities, and government and the TU Program serves as the liaison to establish such collaborations with NASA. NASA technology is an important resource to support the development of new medical products and techniques that will further advance the quality of health care available in the United States and worldwide.

## INTRODUCTION

In the early 1900s, few effective treatments existed for life-threatening diseases because medical care consisted primarily of palliative treatments to mitigate the effects of diseases. Since then, advances in biomedical sciences and technology have produced more effective diagnostic methods and treatments that together have far-reaching effects on medical care and human life expectancy. The pace of technological change in recent decades has been unprecedented, enabling us to reduce infant mortality, to eradicate certain diseases, to replace and repair human organs, and to rehabilitate those with disabilities so that they may live happier and more productive lives.

Through the active transfer of technology, NASA has played an important role in many technological developments that have advanced the quality of medical care in the U.S. and throughout the world. Meeting the aeronautical and space goals of the past three decades has necessitated advancements across a broad spectrum that embraces virtually every scientific and technological discipline. This technology represents an important national resource available for use by public and private sector organizations to develop new or improved products and processes.

To promote the secondary application of aerospace technology in medicine, NASA works with medical device companies, medical universities, major hospitals, and research centers, as well as Federal agencies such as the National Institutes of Health and the Veterans Administration. These collaborations are supported by the NASA (TU) Program whose mandate is to promote the widest practical and appropriate dissemination of the results of NASA's R&D activities. The TU program encourages and facilitates the access and utilization of NASA-developed technology by public and private researchers, developers, and entrepreneurs. The engineers on the NASA Technology Application Team at Research Triangle Institute (RTI) provide an active interface among NASA, research institutions, and private businesses to assist in identifying and developing technology transfer opportunities. Technology Utilization Officers (TUOs) are located at each of NASA's Field Centers and serve as the gateway for external inquiries to access the technological resources of, their respective Centers. The following examples of technology spinoffs illustrate just a few of the benefits realized through secondary applications of aerospace technology.

### Lixiscope

When scientists discovered that some galaxies and stars emit X-rays, they set about developing imaging techniques to enhance the relatively low levels of radiation coming from these far-off sources. Because the new detector developed was able to image low-intensity X-rays, this detector can be used with a much smaller X-ray source for less risky medical imaging procedures. This technology has taken the form of a low-intensity X-ray imaging scope (Lixiscope). Originally developed by Goddard Space Flight Center, the Lixiscope is being produced by Lixi, Inc. of Downers Grove, Illinois. The Lixiscope is a self-contained, battery-powered fluoroscope that produces an instant X-ray image through use of a small amount of radioactive isotope. This unit is designed to utilize less than 1 percent of the radiation required by conventional X-ray devices. Because this system is completely portable, it provides a way to immediately scan athletes for possible bone injuries at the site of the sporting event, or for other such emergency medical uses. Other applications for this system include dentistry and orthopedic surgery. The Lixiscope is finding growing acceptance as an industrial tool in the U.S. and abroad for rapid nondestructive testing for the development of X-ray film. It is also used in security applications, such as examining parcels in mail rooms and building entries.

### Cardiac Defibrillation

Approximately 50,000 people die each year in the U.S. due to ventricular fibrillation, where the electrical conduction system of the heart becomes "short circuited." The muscles of the heart contract in a disorganized fashion, and the heart does not pump blood efficiently. When it can be applied in time, electric shock defibrillation is generally successful in restoring normal electrical activity to the heart. Unfortunately, most of those who die each year from a fibrillation episode are away from a hospital, where they could receive proper treatment.

As the population of astronauts became more diverse, the possibility of a cardiac episode needed to be considered. NASA's Johnson Space Center, in conjunction with International Biomedical of Houston, Texas, developed an advanced defibrillator monitoring system. This system combines the defibrillator paddles with a microprocessor-based monitoring system to display both treatment and patient information. The system is lightweight, portable, and easy to operate. Such equipment is being considered as part of a life support module for future shuttle flights.

In a related project, NASA's Goddard Space Flight Center, Johns Hopkins University, and Intec Systems, Inc., collaborated on the development of an implantable device capable of sensing ventricular fibrillation and delivering a brief defibrillating electrical pulse. The implant incorporates advanced microelectronics and sensors to provide programmability to adapt to individual patient requirements. Subsequently, the ability to detect and correct ventricular tachycardia, another form of arrhythmia, was added to the device. The resulting product, called the Automatic Implantable Cardioverter-Defibrillator (AICD) has been commercialized by Cardiac Pacemakers, Inc. Through telemetry, the device can be interrogated to determine if a defibrillating pulse has been provided, and if so, to look at the electrocardiogram just prior to and following the episode. Hundreds of life saving situations have been documented with use of this system.

### Programmable, Implantable Medication System

Over one million Americans suffer from diabetes, a disease that destroys the body's ability to control its blood sugar and produces such serious complications as heart disease, kidney malfunction, and blindness. Yet most diabetics are able to live long, productive lives by closely monitoring blood sugar levels, continually injecting insulin, and carefully controlling diet and activity levels. Now, a new technology has emerged that may ultimately free diabetics and victims of other long-term diseases from their restrictive lifestyles. Engineers have developed a programmable, implantable medication system (PIMS) that will automatically deliver prescribed doses of medication to key areas in the body, eliminating the need for burdensome self-medication programs. In addition, research has shown that infusion of "short-acting" insulin in small

amounts over a long period - instead of multiple daily injections of "long-acting" insulin - has helped many diabetics achieve better control of blood sugar levels, thereby minimizing the possibility of complications. The device is the size of hockey puck, is encased in a titanium shell, and holds about two-and-a-half teaspoons of concentrated insulin administered at a preprogrammed rate. If a change in measured blood sugar level dictates a different dose, the patient can vary the amount of insulin delivered by holding a small radio transceiver over the implanted system and dialing in a specific program held in the PIMS computer memory. A miniature two-way communications system, based on the space technology of telemetry, sends out signals from the implant with operating information such as insulin usage and pump performance. When an insulin refill is needed, four to six times a year, it is accomplished without surgery by a special hypodermic needle.

Johns Hopkins University's Applied Physics Laboratory (APL) headed the initial development of PIMS as a TU project sponsored by NASA's Goddard Space Flight Center. PIMS is an outstanding example of how space technology offers special utility in medical systems. PIMS employs several technologies derived from R&D work on NASA space systems, including a tiny, microminiaturized fluid control system initially used in life search experiments aboard two NASA Viking spacecraft that landed on Mars. MiniMed Technologies of Sylmar, California, licensee of the technology, has been refining the design of PIMS since the initial development at APL. The PIMS unit has performed well in clinical trials and is entering a second phase of trials for various applications. The technologies in PIMS have also been incorporated in portable external pumps marketed by MiniMed and by the Biomedical Group of Parker Hannifan Corporation, of Irvine, California. These pocket-sized micropumps allow ambulatory infusion of chemotherapeutic, antibiotic, and antipain medications.

### On the Horizon

Two human imperatives, exploration and improving the condition of mankind, have found common ground in the space program. Technologies developed to meet the awesome challenges of space exploration in the past three decades have paid extra dividends in improving the quality of life on Earth. Medical innovations derived from space technology that may be seen in the next decade include: a technique for noninvasive measurement of intracranial pressure, a critical parameter in the treatment of head injury; a device for management of dangerous wandering behavior in Alzheimer's patients; improved heart valves made possible by computational fluid dynamics; an analytical technology developed for aerospace applications; and an adjustable shunt for improved control of cerebrospinal fluid pressure in children with hydrocephalus.

## Direction Discriminating Hearing Aid System

M. Jhabvala  
NASA/GSFC, Greenbelt, MD  
H.C. Lin  
University of Maryland, College Park, MD  
G. Ward  
DTI, Engineering, Oxon Hill, MD

### Abstract

A visual display system has been developed for people with substantial hearing loss in either one or both ears. The system consists of three discreet units: an eyeglass assembly for the visual display of the origin or direction of sounds; a stationary general purpose noise alarm; and a noise seeker wand.

### Introduction

Individuals who suffer a severe hearing loss in one ear have great difficulty discerning the direction from which sounds originate. Situations exist where individuals must be made immediately aware of the origin of the sound, especially sound intended as a warning. Typical examples include: street traffic, loading dock horns, sirens and a variety of other less common audibles but still of significant warning authority. Other situations exist where hearing impaired individuals need to be cognizant of an audible signal. A stationary alarm was developed to assist individuals in situations such as a knocking on a door, a baby crying in another room or even a faucet left running. In conjunction with the stationary alarm a noise seeker was also developed to assist individuals in locating the origin of the sound if not readily apparant. The Solid State Device Development Branch (Code 724) supported by the GSFC Office of Commercial Programs embarked on a program to develop systems which would address these needs. The University of Maryland's Electrical Engineering Department provided design and fabrication expertise and after successful breadboard tests Design Three Incorporated developed a commercial prototype under NASA contract.

### Method

#### Eyeglass Assembly

The most complicated of the three units is the eyeglass assembly. Aside from performing the necessary decision making and display functions the electronics had to fit compactly into an eyeglass frame. The components in the eyeglass include a custom developed CMOS integrated circuit, two miniature batteries, two miniature sensing microphones, capacitors, resistors, potentiometers and light emitting diodes (LEDs).

The electrical organization of the system is shown in Fig.1. An audio signal is detected by each of the two microphones positioned in the eyeglass stems. Keeping in mind that the direction from which the sound originated is the desired parameter the signal generated by the microphone more closely facing the audio source will be the stronger of the two. Both microphone signals are amplified and then rectified. This rectified d.c. signal is then processed through an 3-bit analog to digital (A/D) converter. The two digital signals are encoded to binary numbers identifying the signal in one of the three volume level categories. The signals are compared and only the stronger signal is permitted to travel forward. There are seven LEDs, three on the top of each eyeglass lens and one at the bottom of one of the lenses. A red, green or yellow LED on the lens will illuminate based on the loudness of the sound received on that side of the individual. Based on the binary encoding, logic circuitry will determine which of the three LED's should be illuminated, red (loud), green (medium) or yellow (soft). In the instance where the sound originated directly in front of or behind the individual an equal LED (clear) is illuminated indicating a strong enough sound but either directly in front of or directly behind the individual.

In order to compact all the electronics into a minimal area a custom integrated circuit was designed and manufactured. All the electronics including the A/D, encoder, comparator, logic and LED drivers were condensed to an area 0.18 inches by 0.26 inches. The chip is powered by two lithium batteries ( $\pm 3v$ ). Fig. 2 is a photograph of the integrated circuit. A schematic layout of the eyeglass assembly is shown in Fig. 3.

### Stationary Alarm

The stationary alarm is a free standing portable unit with a high intensity xenon flash tube as the light indicator. It can be placed near a door, for example, to visually alert an individual that there is a knock or door bell ring occurring. An external control provides signal adjustment to control the xenon light threshold flash level and the unit as currently designed operates from a 110 volt supply. The stationary alarm unit is shown in Fig. 4.

### Noise Seeker Unit

The third component of this ensemble is the noise seeker unit. This unit has LED bar graph displays as the indicator and is powered by rechargeable batteries. This unit is intended to be used in conjunction with the stationary alarm to assist in locating the origin of a noise signal that triggered the alarm. The noise seeker plugs into the stationary alarm where it is recharged when the stationary alarm unit is powered by 110 AC. This feature also keeps the noise seeker unit in close proximity to the alarm. The seeker uses highly uni-directional microphones. When the unit is aimed directly at the noise source the bar graph LED's will illuminate. As the wand is rotated away from the noise source fewer and fewer LED's will illuminate thus giving the bearer a good indication in which direction to proceed. A diagram of the noise seeker unit is shown in Fig. 5.

### Conclusion

In an effort to transfer some of the microelectronics and miniaturization techniques used at the NASA/Goddard Space Flight Center to the public sector a system has been conceived, designed, developed and now commercially prototyped to assist hearing impaired individuals. The work, sponsored by Mr. Don Friedman of the Office of Commercial Program, was a collaborative effort involving numerous graduate students at the University of Maryland, other government agencies, private institutions and industry. It is hoped that NASA and the governments role will have been successfully completed and that these units will ultimately find their way to those individuals who need them.

### References

Jhabvala, M.; Lin, H.C.; Wang, Z.S; Chen, T.; and Zong, Q.:  
Direction Discriminating Hearing Aid System  
RESNA 12th Annual Conference; New Orleans, 1989.

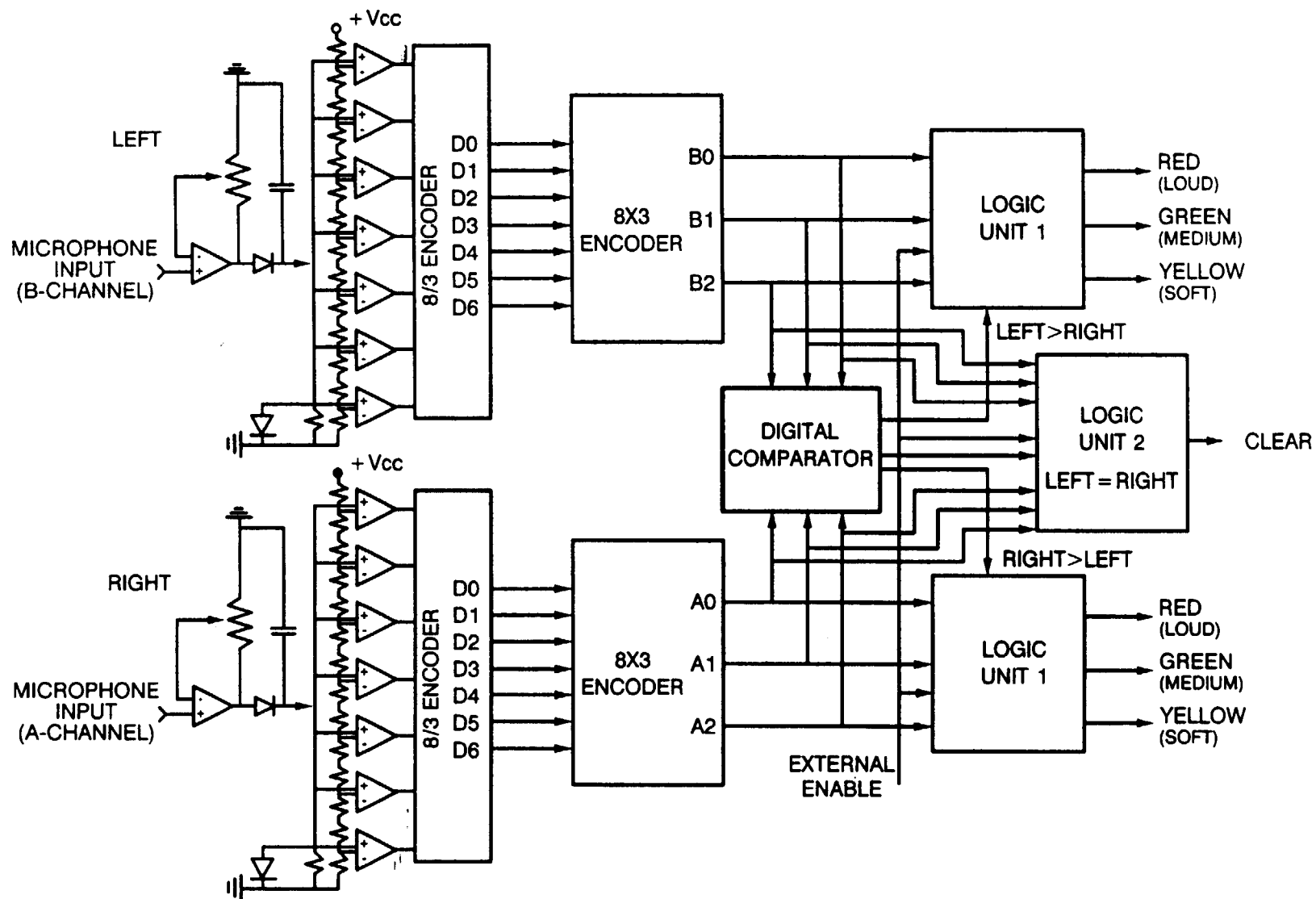


FIGURE 1 BLOCK DIAGRAM OF THE SYSTEM ELECTRONICS FOR THE EYEGLOSS ASSEMBLY



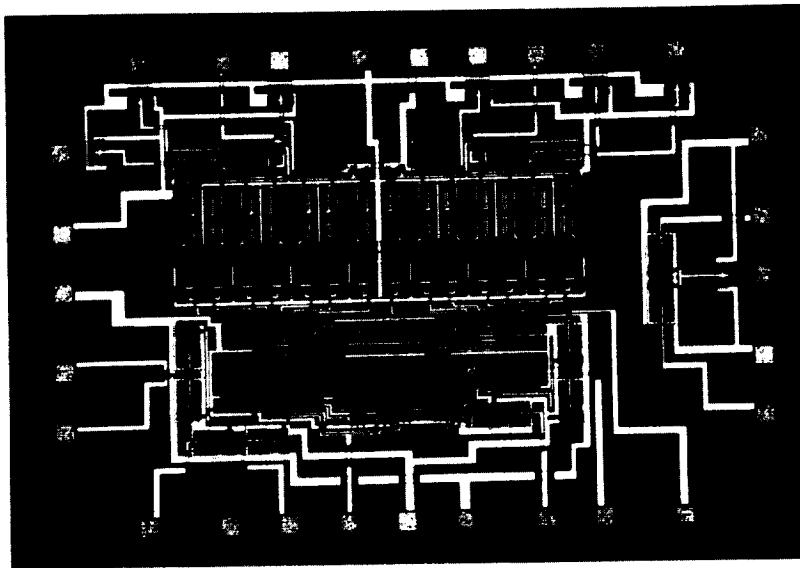


FIGURE 2 PHOTOMICROGRAPH OF CMOS INTEGRATED CIRCUIT

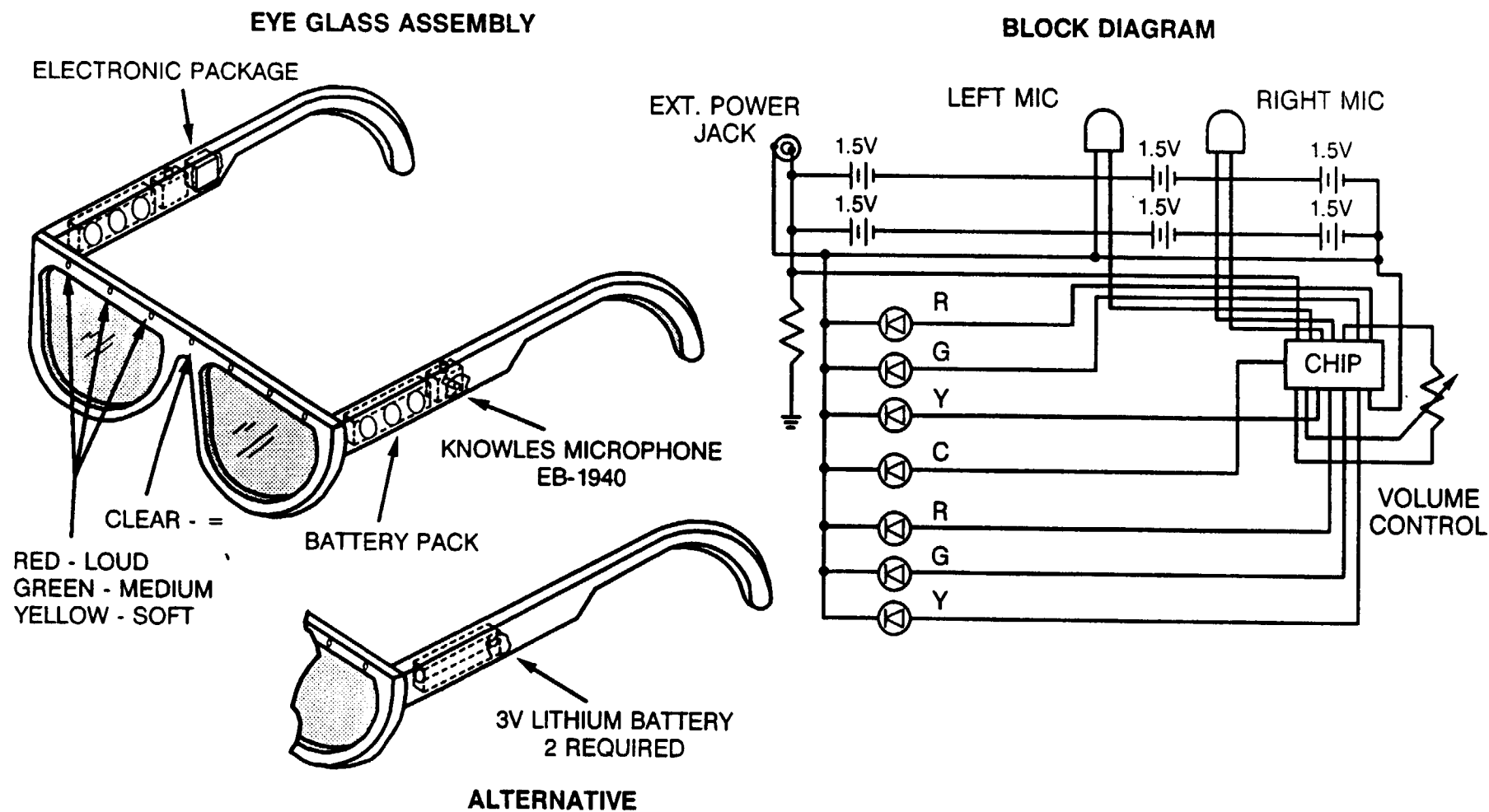


FIGURE 3 EYEGLOSS ASSEMBLY AND LED DISPLAY DETAIL

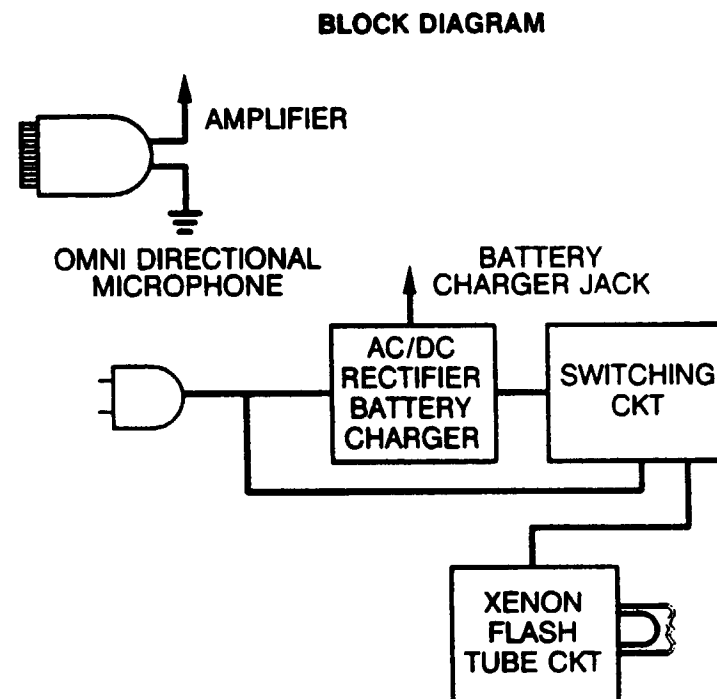
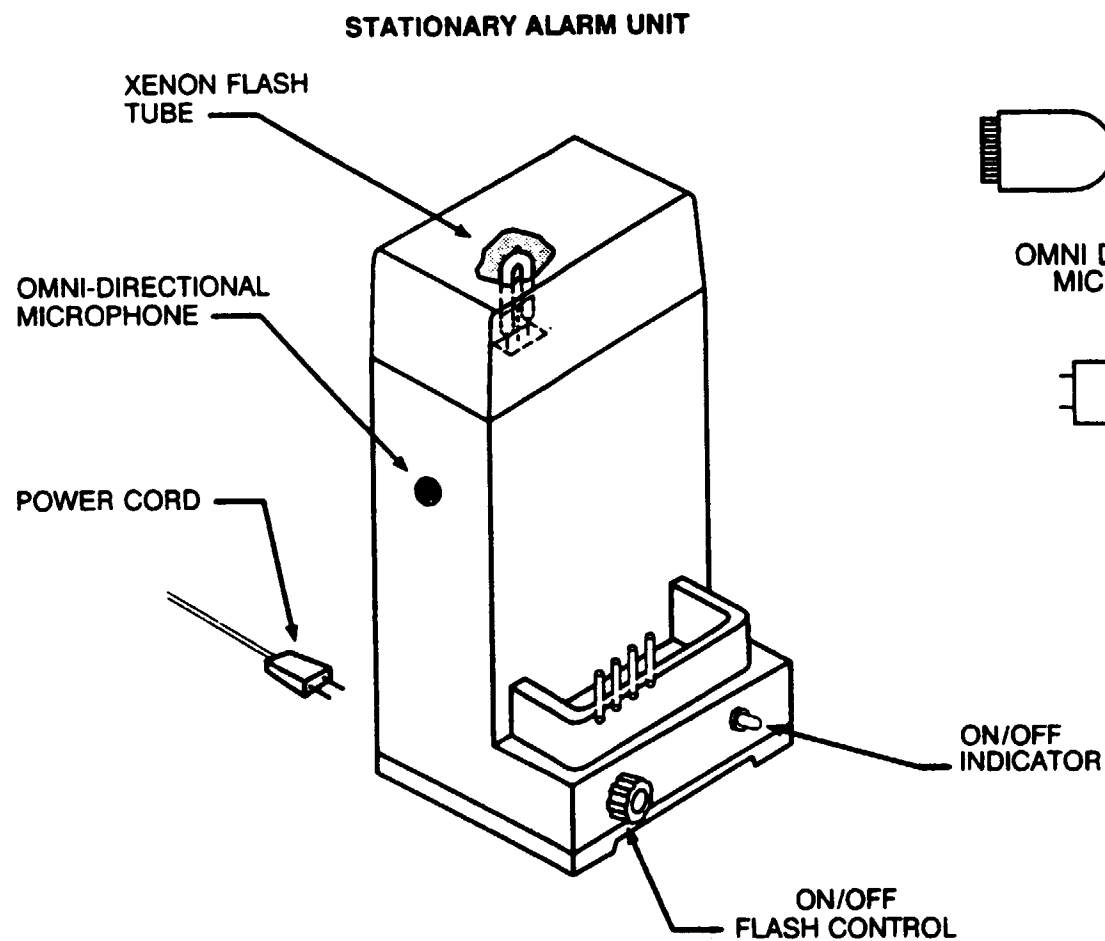


FIGURE 4 STATIONARY ALARM UNIT

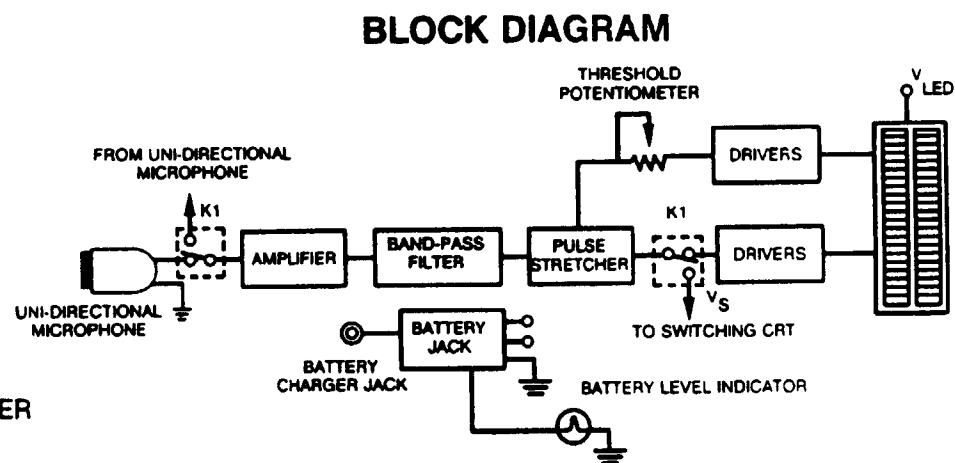
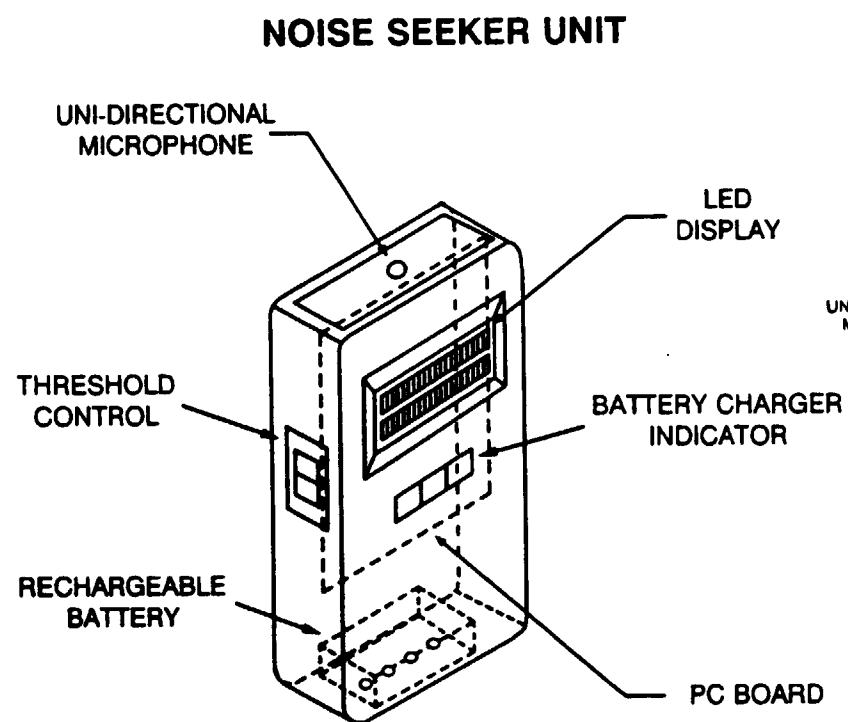


FIGURE 5 NOISE SEEKER UNIT AND SYSTEM BLOCK DIAGRAM

## X-RAY IMAGING MICROSCOPE FOR CANCER RESEARCH

**Richard B. Hoover**Space Science Laboratory, ES-52  
Marshall Space Flight Center, AL 35812**David L. Shealy**Dept. of Physics  
University of Alabama at Birmingham  
Birmingham, AL 35294**B. R. Brinkley**Dept. of Cell Biology and Anatomy  
University of Alabama at Birmingham  
Birmingham, AL 35294**Phillip C. Baker**President, Baker Consulting  
Walnut Creek, CA 94596**Troy W. Barbee, Jr.**Lawrence Livermore National Laboratory,  
Livermore, CA 94550**Arthur B. C. Walker, Jr.**Center for Space Science and Astrophysics, Stanford University,  
Stanford, CA 94305

## ABSTRACT

The NASA technology employed during the Stanford/MSFC/LLNL Rocket X-Ray Spectroheliograph flight established that doubly reflecting, normal incidence multilayer optics can be designed, fabricated and used for high resolution x-ray imaging of the Sun. Technology developed as part of the MSFC X-Ray Microscope program, demonstrated that high quality, high resolution multilayer x-ray imaging microscopes are feasible. Using technology developed at Stanford University and at the DOE/Lawrence Livermore National Laboratory (LLNL), Troy W. Barbee, Jr. has fabricated multilayer coatings with near theoretical reflectivities and perfect bandpass matching for a new rocket-borne solar observatory, the Multi-Spectral Solar Telescope Array (MSSTA). Advanced Flow Polishing has provided multilayer mirror substrates with sub-angstrom (rms) smoothness for our astronomical x-ray telescopes and x-ray microscopes. The combination of these important technological advancements has paved the way for the development of a *Water Window Imaging X-Ray Microscope* for cancer research.

This instrument is a doubly reflecting multilayer coated x-ray microscope configured to operate within the "water window," a narrow regime of the x-ray spectrum between the K absorption edges of oxygen (23.3 Å) and of carbon (43.62 Å). In this wavelength regime, water is relatively highly transmissive and carbon is highly absorptive, permitting the microscope to delineate carbon based structures within living cells. The development of this high spatial resolution and high contrast capacity to image living cells, in aqueous physiological environments, will afford advantages not available in any conventional microscopes. The Water Window Imaging X-Ray Microscope offers non-invasive strategies for examining living tumor cells without the need of dyes, stains or exogeneous chemicals which produce limiting artifacts. Our theoretical analysis has shown that multilayer x-ray microscopes of the Schwarzschild configuration should achieve spatial resolution in the 100 Å range or better. Such performance could permit direct imagery of cytoskeletal components, membranes, secretory vesicles, endoplasmic reticulum, chromatin, nucleoli and nucleosomes. It should improve diagnosis and greatly benefit experimental studies of tumor cell biology. Advanced versions using aspheric optics and multiple elements may achieve spatial resolution sufficient to resolve DNA and RNA molecules within living tumor cells (the double helix of the DNA molecule is typically only 20 Å wide, but very long). In this paper, we will describe the design of the Water Window Imaging X-Ray Microscope and discuss technological aspects of mirror fabrication, optical assembly, alignment and testing of the instrument. We predict the optical performance and consider applicability of this microscope to studies of cell biology in general and tumors in particular.

## INTRODUCTION

On Oct. 23, 1987 the Stanford/MSFC/LLNL Rocket X-Ray Spectroheliograph<sup>1</sup> was successfully launched on a Nike-boosted Black Brant sounding rocket from the White Sands Missile Range, New Mexico. This flight produced the first high resolution, narrow wavelength band x-ray/EUV images of the Sun (Fig. 1) obtained with normal incidence multilayer x-ray optics. We had previously produced high resolution x-ray images at 44 Å with doubly reflecting multilayer telescopes at the MSFC X-Ray Calibration Facility.<sup>2</sup> These laboratory and solar images constitute dramatic proof that normal incidence multilayer coatings on curved surfaces can produce superb x-ray/EUV images. Photons of wavelengths of 1-100 angstroms (Å) are considered to be x-rays, and 100 - 1000 Å comprise the Extreme Ultraviolet (EUV) regime. The high density solar images we recorded on high resolution, but relatively insensitive, photographic emulsions demonstrated that multilayer coatings can yield excellent x-ray/EUV reflectivities and that the bandpasses of peak reflectivity of the two mirrors of a doubly reflecting multilayer imaging system can be precisely matched.

ORIGINAL PAGE

BLACK AND WHITE PHOTOGRAPH

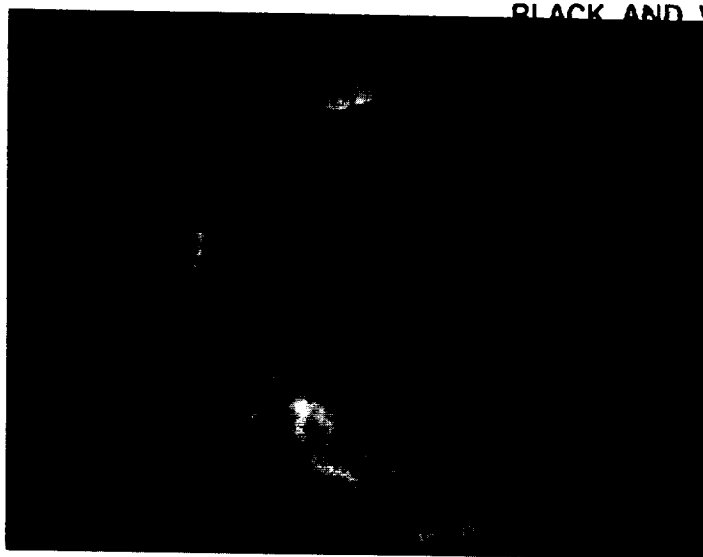
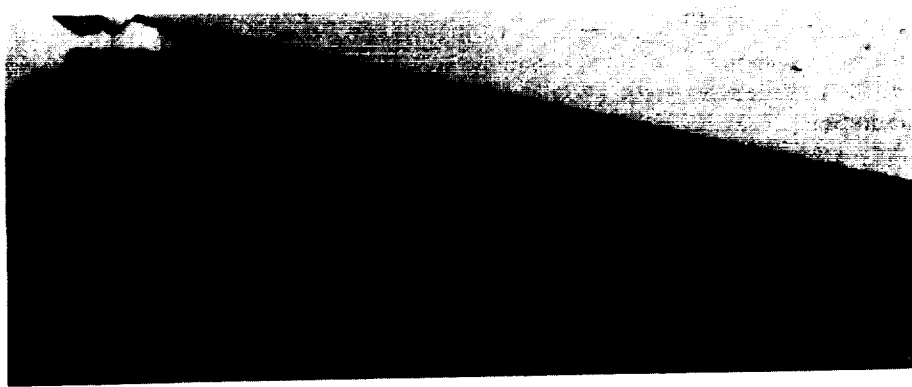


Fig. 1. High resolution 173 Å image of the Sun produced with a doubly reflecting Cassegrain multilayer x-ray telescope operating at normal incidence.

Normal incidence multilayer x-ray optics are the result of the pioneering work of one of the authors, Troy W. Barbee, Jr. at Stanford University<sup>3</sup> and Eberhard Spiller<sup>4</sup> of the IBM Watson Research Center. Their revolutionary breakthroughs in materials science technology, which permitted the fabrication of stable multilayer x-ray optical coatings, were independently achieved only a little over a decade ago. More recently, while at the Lawrence Livermore National Laboratory (LLNL), Barbee has fabricated multilayer coatings, of unprecedented quality, with near theoretical reflectivities at x-ray/EUV wavelengths. Precise bandpass matching was achieved for the primary and secondary optical elements of telescopes which will be flown by NASA. This payload, known as the Multi-Spectral Solar Telescope Array (MSSTA),<sup>5</sup> is larger than its predecessor and will be launched on a Terrier-boosted Black Brant sounding rocket early in 1991. The 127 mm aperture MSSTA telescopes, each with four times the collecting area of the Cassegrain telescope previously flown, employ Ritchey-Chrétien optical systems (Fig. 2). In these telescopes, the primary and secondary mirrors are hyperboloidal. These are "aplanatic" telescopes, which means that the optical aberration known as coma is zero (to third order at least) and spherical aberration is absent. This design permits the telescopes to produce high spatial resolution images over a wide field of view. Theoretical calculations and laboratory studies of these instruments indicate that these telescopes should produce full-disk solar x-ray/EUV images with spatial resolution as high as 0.1 arc second.<sup>6</sup>

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



**Fig. 2.** Completed MSSTA Ritchey-Chrétien multilayer telescopes, using fabrication technology similar to that planned for the Water Window Imaging X-Ray Microscope.

During the development of the MSSTA telescopes, it was also established that the application of the Advanced Flow Polishing technology (pioneered by one of the authors, Phillip C. Baker of Baker Consulting) can yield mirror substrates with sub-angstrom level rms surface smoothness.<sup>7</sup> Based upon these results, a program was initiated at MSFC under the auspices of the Center Director's Discretionary Fund (CDDF) to develop normal incidence magnifying Schwarzschild x-ray microscope optics for coupling X-Ray telescopes to detectors for resolution enhancement. The microscope optics were produced on Zerodur blanks to smoothness of 2-3 Å rms. However, during this program, it was also shown that contoured surfaces with unprecedented rms smoothness in the range of 0.5-0.6 Å could be produced on Hemlite grade Sapphire blanks by Advanced Flow Polishing.

These profoundly important technological advancements in the ability to fabricate high reflectivity, matched bandpass, normal incidence optics operating at 44 Å and the ability to produce the ultra-smooth substrates required by these coatings led us to conclude that a Water Window Imaging X-Ray Microscope is feasible. This instrument affords great promise as a fundamental tool for basic cell biology and cancer research.

### MULTILAYER X-RAY OPTICS

Multilayer x-ray mirrors are essentially synthetic Bragg crystals that can be contoured to a figured surface. (For a detailed description of Multilayer X-Ray Mirrors see Barbee<sup>8</sup>). They are fabricated by the accurate deposition on an ultra-smooth substrate of a coating consisting of a stack of many alternating layers of high atomic number (high-Z) diffractor material separated by layers of a low-Z spacer material. The layers must be very uniform and of precisely repeatable thicknesses  $d_1$  and  $d_2$ , respectively. Since this multilayer coating constitutes a synthetic Bragg crystal, x-ray reflection occurs by the process of Bragg diffraction. When slight refraction effects are ignored, the wavelength at which the peak of the reflectivity occurs is given by the Bragg relation:  $n(\lambda) = 2D \sin(\theta)$ , where  $D = d_1 + d_2$  and  $\theta$  is the angle at which the radiation strikes the mirror, as measured from the mirror surface. For mirrors operating at normal incidence ( $\theta = 90^\circ$ ), the equation becomes:  $\lambda = 2D$ , where  $\lambda$  is the wavelength of peak reflectivity of the first order Bragg diffracted light. Since  $D$  is the sum of the two layer thicknesses, a multilayer coating, designed to reflect 44 Å x-rays, could be produced as a stack of alternating 11 Å thick layers of a high-Z material (such as tungsten carbide) separated by 11 Å thick layers of a low-Z spacer material (such as carbon). Considering the dimensions involved, it is clear that these layers are only a few atoms thick. Furthermore, the diffracting layers must be coplanar and uniform. Since the layers follow the contour of the substrate upon which they are deposited, there are very stringent requirements on the smoothness and uniformity of multilayer optics substrates. This is especially important for layers with thicknesses in the 6 - 11 Å regime, which are necessary for systems designed to operate at normal incidence in the water window. For the multilayer coating to be an effective reflector, there must be many layers in the stack and all must be of the same thickness to a very high degree.

of accuracy. Indeed, by choice<sup>9</sup> of the D spacing and the materials comprising the multilayer coating, it is possible to tailor the coating to reflect very narrow bandpasses at selected wavelengths of x-ray/EUV radiation.

Although only a small fraction of the incident radiation is reflected at each low-Z/high-Z interface, by use of a stack of tens to hundreds of alternating layers in the coating, high reflectivities at normal incidence can be achieved by constructive interference if the layer pairs are deposited with sufficient uniformity. Recently, tests were performed at the Stanford Synchrotron Radiation Laboratory (SSRL) and the National Institute of Standards and Technology (NIST) SURF II Synchrotron on the MSSTA telescopes. These studies revealed that reflectivities approaching 45% at normal incidence were produced by the mirrors operating at 173 Å and 193 Å. For our solar rocket program, optics were fabricated with coatings on both convex and concave superpolished substrates whose reflectivities were peaked for x-ray and EUV radiation in the range  $44 \text{ Å} < \lambda < 335 \text{ Å}$ .

Since multilayer x-ray mirrors reflect x-rays by the phenomenon of Bragg diffraction, only a very narrow bandpass is efficiently reflected wherein the Bragg condition is satisfied. In the water window, multilayer mirrors can achieve spectral resolution ( $\lambda/\Delta\lambda$ ) exceeding 50. This characteristic is of the utmost importance for the development of a *Water Window Imaging X-Ray Microscope*. X-rays of longer or shorter wavelength would seriously degrade the contrast of carbon structures within the cell, which is a major reason why grazing incidence x-ray optics are not well suited as optics for the fabrication of a Water Window X-Ray Microscope. Grazing incidence optics reflect x-rays over a much broader bandpass than the water window. They are also far more sensitive to contaminants and x-ray scattering and suffer more severely from optical aberrations than normal incidence multilayer optics. Because multilayer optics reflect in only a very narrow bandpass of the incident radiation, precise matching of the wavelength at which peak reflectivity occurs from the primary and secondary mirrors is required. If these bandpasses are not accurately matched the net throughput of the instrument will be drastically reduced.

## X-RAY IMAGING IN THE WATER WINDOW

The biological significance of the x-ray water window is due to the fact that water is prevalent in all living cells. The oxygen of the water plays the dominant role for the absorption of soft x-rays. However, the structures that are of the greatest scientific interest (such as organelles, cytoskeletal components, membranes, secretory vesicles, endoplasmic reticulum, chromatin, nucleoli and nucleosomes) are in general comprised of complex molecules (DNA, RNA, proteins, etc.) incorporating large amounts of carbon. The nature of the interaction of x-rays with matter makes it possible to observe these carbon structures with minimal interference from the surrounding water. A sharp discontinuity or "edge" in the absorption spectrum of a material occurs when the energy of the photon is sufficient to ionize electrons from one of the shells or sub-shells of the atoms comprising the material. These edges are designated by the shell or subshell from which the electrons are ejected, i.e. K for the innermost shell,  $L_I$ ,  $L_{II}$ , and  $L_{III}$  for the sublevels of the next shell, etc. The strongest *absorption edges* are the K edges. For wavelengths immediately below the K edge, the absorptivity increases dramatically. Since the x-rays whose wavelength are longward of the K absorption edge do not have sufficient energy to eject a K-shell electron, they are not strongly absorbed by the material, which appears relatively transparent at these wavelengths.

For the purposes of x-ray microscopy, it is significant that the K absorption edge for the element oxygen lies at 23.3 Å and for carbon the K edge is at 43.62 Å. This results in a narrow bandpass in the soft x-ray spectrum between 23.3 Å (oxygen K absorption edge) and 43.62 Å (carbon K absorption edge) called the "water window." In this wavelength regime, water is relatively transparent and carbon is highly absorptive. The opacity of protein and water at these wavelengths has been calculated by London et al.<sup>10</sup> and are shown in Fig. 3. These results show the dramatic difference in the absorptivity of protein and water within the water window. The Water Window Imaging X-Ray Microscope will make it possible to investigate carbon structures (and possibly even the motions of those structures) within the aqueous environment of living cells.

We have, therefore, initiated a program to fabricate an imaging x-ray microscope, utilizing ultra-smooth, Advanced Flow Polished and figured sapphire mirror substrates coated with multilayers,



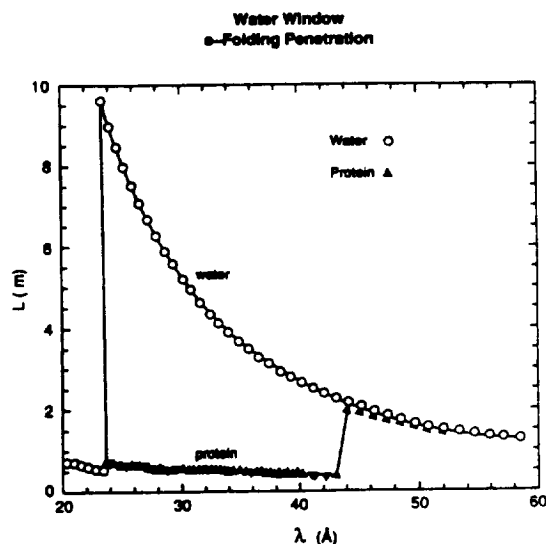


TABLE I. Schwarzschild mirror parameters.

$M(x)$	$R_1$ (cm)	$R_2$ (cm)	$s$ (cm)	$d$ (cm)	$z$ (cm)	$f$ (cm)
2	105.00	10.00	18.25	98.00	-91.42	5.51
3	58.27	10.00	18.05	48.27	-34.12	6.04
4	45.58	10.00	18.01	35.58	-13.56	6.41
5	40.00	10.00	18.00	30.00	0.0	6.67
10	31.79	10.00	18.02	21.79	48.36	7.29
15	29.69	10.00	18.04	19.69	91.04	7.54
20	28.75	10.00	18.05	18.75	131.49	7.67
30	27.84	10.00	18.06	17.84	213.27	7.80
40	27.10	10.00	18.07	17.40	296.07	7.87
50	27.15	10.00	18.07	17.15	376.42	7.92

Fig. 3. E-Folding penetration of x-rays in the water window. Re-drawn from London *et al.*<sup>10</sup>

with 2D spacing such that  $23.3\text{\AA} < 2D < 43.62\text{\AA}$ , as is appropriate to reflect x-rays of a narrow bandpass within the water window. At these wavelengths, ultra-high resolution photographic films (i.e. XUV 100 and XUV 649) and photo-resists can be used as the x-ray detector. Since the 649 emulsion affords spatial resolution of 2000 lines/mm when processed for optimum resolution, and a very great dynamic range, it will serve as a primary detector for the instrument. This microscope should be capable of producing high resolution, high contrast images of chromosomes, proteins and other carbon structures within living or freshly killed cells. The *Water Window Imaging X-Ray Microscope* should permit smaller structures to be resolved than is currently possible with visible light or fluorescence microscopy. By obtaining sequential images using high repetition rate laser plasma x-ray sources, it may be possible to investigate motions of genetic material, proteins and other structures within living cells.

### OPTICAL CONFIGURATION OF THE MULTILAYER X-RAY MICROSCOPE

The exciting results which we obtained with normal incidence multilayer x-ray telescopes encouraged us to continue our efforts to develop an aplanatic imaging x-ray microscope utilizing multilayer optics in the Schwarzschild configuration. Normal incidence multilayer x-ray optics for use in scanning x-ray microscopes were studied by Spiller<sup>11</sup> using elliptical mirrors. Trail and Byer<sup>12</sup> fabricated a scanning microscope and Lovas *et al.*<sup>13</sup> employed a Schwarzschild system for laser fusion research. Our prior studies of Schwarzschild multilayer microscopes were constrained to the development of systems for which high smoothness spherical laser mirrors were available as "off-the-shelf" components. Suitable spherical substrates had been purchased from General Optics of Moorepark, CA, and were used for the telescopes flown on Oct. 23, 1987. However, during the development of the MSSTA, Baker Consulting produced hyperboloidal optical substrates of ultra-high smoothness (1-3 Å rms). His fabrication methods can also yield spherical or aspheric substrates which are ideal for x-ray microscopes. High resolution aplanatic imaging x-ray microscopes configured from low x-ray scatter normal incidence multilayer optics should find important applications in many areas, including laser fusion research, x-ray lithography, materials science, astronomy, genetic engineering, virology and bacteriology, as well as fundamental cell biology and cancer research.

We have designed and analyzed several Schwarzschild x-ray microscope configurations. Diffraction analysis indicates better than 200 Å spatial resolution in the object plane for up to a 1 mm field of view can be achieved with 125 Å radiation. Since the diffraction limit scales with the wavelength, when the microscope is used with 37 Å radiation (within the water window) spatial resolution well below 100 Å may be realized. We are currently fabricating 20X and 30X normal incidence multilayer x-ray microscopes of 1.35 meter overall length. An aplanatic x-ray microscope using two spherical mirrors can be constructed by imposing the Schwarzschild condition on the selection of the mirror radii. The Schwarzschild condition can best be understood by referring to Fig. 4.

The mirror surfaces  $S_1$  and  $S_2$  are concentric spherical surfaces of radii  $R_1$  and  $R_2$ , respectively. A complete discussion of the ray trace analysis of a Schwarzschild microscope configured for normal incidence multilayer applications was presented by Hoover et al.<sup>14</sup> and Shealy et al.<sup>15</sup>. The Schwarzschild condition for an aplanatic, two mirror imaging system can be expressed:

$$\frac{R_2}{R_1} = 1.5 - \frac{R_2}{Z_o} \pm \left[ 1.25 - \frac{R_2}{Z_o} \right]^{1/2}, \quad (1)$$

where the "+" sign is used in Eq. 1 for magnifications greater than 5, and the "-" sign is used for magnifications less than 5. Hoover et al.<sup>14</sup> have summarized the Schwarzschild design equations and presented the dependence of the rms blur circle radius as a function of the object height, image plane location, mirror tilts, and decentration for a 10x microscope with a total length of 1.41 m. As the magnification increases so does the overall length of the microscope. The parameters for systems varying in magnifications from 2X to 50X have been computed and are given in Table 1. The mirror substrates which we have fabricated were selected for 20X and 30X Schwarzschild systems.

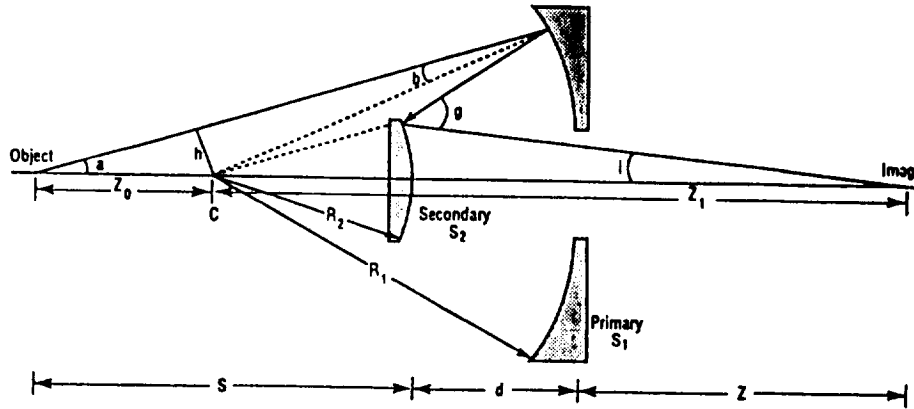


Fig. 4. Schwarzschild configuration for an aplanatic normal incidence x-ray microscope.

We have calculated the spatial resolution, transmission losses due to vignetting, and off-axis performance for these microscopes. These calculations imply that in order to take advantage of this high spatial resolution, photographic emulsions capable of achieving 0.78 micron spatial resolution will be required. This implies the need for films capable of resolving better than 1300 lines per mm over a 20 mm diameter regime. We have established that the Kodak 649 emulsion has the required spatial resolution and is sensitive over the soft x-ray/EUV portions of the spectrum.<sup>16</sup>

Even higher resolution may be achieved by the use of aspheric optics. We have carried out a theoretical design and analysis of aspheric x-ray microscope configurations, which yield far better resolution over a wider field of view than is possible with the Schwarzschild configuration<sup>17</sup>. Results of the analysis of the the aspherical 20X microscope also reveal that superior off-axis performance due to reductions in coma can be achieved.

## FABRICATION OF THE X-RAY MICROSCOPE

The 20X and 30X Schwarzschild x-ray microscopes are now being fabricated utilizing much of the technology implemented in the MSSTA program. Baker Consulting has fabricated mirror substrates from both Zerodur and Hemlite grade Sapphire. It is very important that the mirror substrate material have the ability to be polished to an ultra-smooth finish, but also have low thermal expansion coefficient. Although sapphire has somewhat higher thermal expansion properties than Zerodur, it can be polished to phenomenal smoothness. In addition the sub-surface condition of the mirror substrate must be considered as a possible factor in the performance due to possible stress relaxation during coating or from externally applied force either thermal or mechanical. It has

been demonstrated, during the fabrication of sapphire surfaces, that the use of the Advanced Flow Polishing technique has produced a zero sub-surface damage condition as measured with Rutherford backscatter techniques by General Ionex Corp.

The primary mirrors were fabricated as concave spheres of 8 cm outside diameter. They have a radius of curvature of 23 cm and a central hole diameter of 2.2 cm. The convex spherical secondaries are of 2 cm diameter with an 8 cm radius of curvature. The Sapphire optical surfaces were polished to 0.5-0.6 Å rms surface smoothness and surface figure accuracy better than  $\lambda/10$  when tested with visible light. The mirror substrate smoothness was measured using a Zygo profilometer and the completed optical systems were tested by interferometric techniques at visible wavelengths to insure that the precise optical figure of the elements was obtained. Final performance testing of the completed x-ray microscope assembly will be carried out by producing images of microscopic structures with the instrument at x-ray wavelengths. These tests require the use of ultra-high resolution photographic emulsions and high intensity x-ray sources. Initial studies will be carried out using the laser fusion plasma produced by the OMEGA Facility of the University of Rochester, and with x-ray/EUV emission generated by the NIST SURF II or the Stanford SSRL synchrotron facilities. All microscope structures have been designed and are currently being fabricated. The mount structures are of stainless steel, with a conical configuration so as to be compatible with the beam constraints of the OMEGA Facility. The x-ray microscope currently being fabricated is shown in Fig. 5.

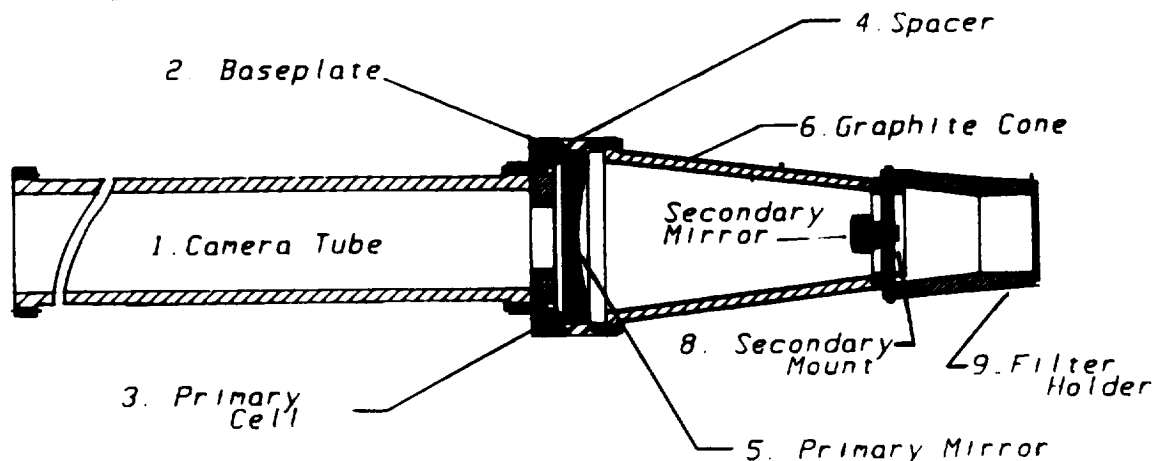


Fig. 5. Structural configuration of the *Water Window Imaging X-Ray Microscope*.

The microscope tube structures are being fabricated by filament winding methods using AS4-12K graphite fiber with an HBRF55A epoxy resin matrix. Longitudinal fibers are applied to increase stiffness and to produce microscope tube structures with near zero coefficient of thermal expansion. This is the same technique used to fabricate the tube structures for the 127 mm diameter telescopes for MSSTA. Thin film filters of 1500 Å aluminum on a nickel support mesh have been fabricated by the Luxel Corp. of Friday Harbor, Washington. These will be used as both specimen supports and visible light rejection filters. The microscope camera adapters accommodate either the 35 mm Canon T-70's, which were flown on the Rocket X-Ray Spectroheliograph, or the 70 mm Pentax 645 MSSTA cameras. Primary data recording will be on photographic film, utilizing experimental XUV-100, 101-07, and XUV 649 emulsions.

### MEASURED PERFORMANCE OF X-RAY MICROSCOPE OPTICS

We have fabricated the optical components for 20X and 30X Schwarzschild x-ray microscopes using Zerodur and sapphire. The mirrors were flow polished, as noted above, and a surface smoothness of 1 Å rms was achieved on Zerodur and 0.5-0.6 Å on sapphire. The surface figure achieved on the mirrors was far better than the  $\lambda/10$  specification. In fact, the completed optical systems were tested interferometrically and found to have superb performance characteristics, with rms wavefront

errors less than  $\lambda/100$ . Figure 6a. shows the measured geometric radial energy distribution of the a 20X Zerodur Schwarzschild mirror system. It can be seen that 90% of the energy is contained within 0.12 Airy disk radii. The superb point spread function of this system, as determined by interferometric analysis of the optics, is shown in Figure 6b.

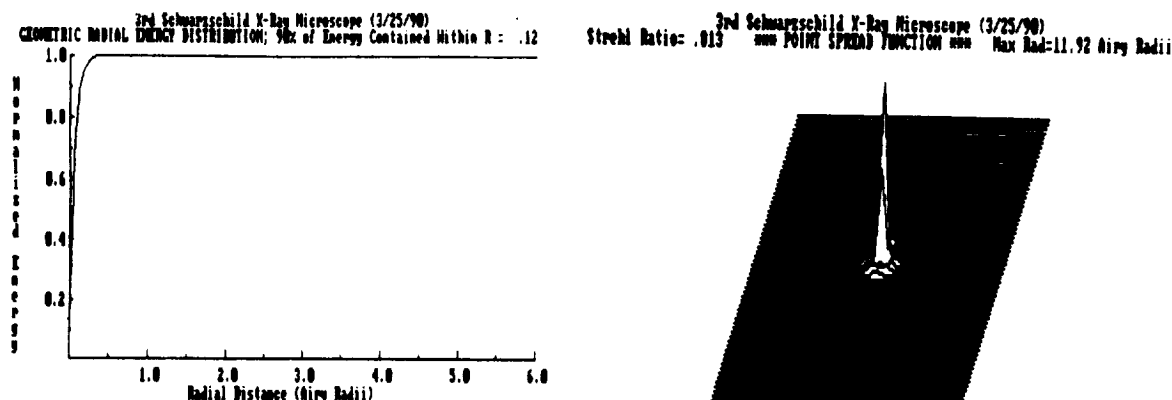


Fig. 6. Multi-pass interferometer measurements of (a.) the geometric radial energy distribution and (b.) the Point Spread Function for the 20X Schwarzschild microscope.

The first Schwarzschild microscope prototype x-ray optics are being coated with multilayers for operation at 130-135 Å. Normal incidence reflectivities better than 55% have been experimentally achieved from multilayers on concave surfaces in this wavelength range. Based upon the theoretical analysis and measurements of the optical characteristics of the completed mirrors which we have performed, we anticipate a spatial resolution of 300-400 Å in the object plane should be obtainable with a Schwarzschild microscope operating at an initial test wavelength of 130-135 Å. In the water window with mirrors of the same design coated to reflect radiation of 36-40 Å wavelength, significantly better spatial resolution should be realized. After final assembly and optical alignment of the multilayer microscope at MSFC, x-ray tests and utilization of the instrument for imaging applications will begin at SSRL, SURF II and the OMEGA Facility. The imaging and assessment phase at synchrotrons and the laser fusion facility could be accomplished within a year.

The requirements for the performance of the optical systems have become more exacting, driven by the high quality of the multilayer coatings and the demands of the imaging requirements at the shorter wavelengths. This has led to our use of special optical testing techniques for the measurement of the surfaces and the contours of the microscope and telescope optics. The standard interferometric analysis at the longer wavelengths (i.e., 6328 Å Helium-Neon line) has limitations especially when the wavelength of use is considerably shorter, such as is the case of the water-window microscope systems. The use of higher sensitivity interferometric testing techniques, or the construction of an x-ray interferometer utilizing multilayer optics is becoming more essential due to the direct effect that system wavefront errors have on imaging quality. Existing techniques, such as multi-pass interferometry and holographic interferometry, have been used to increase the sensitivity of wavefront error detection.

These techniques must be improved as the wavelength shortens, especially in the area of aspheric system testing. The current designs that have been used involve spheres, but we are being considering systems employing aspheric mirrors with both moderate and severe aspheric departures, with unusual contours that are not easily tested. Multi-pass interferometry was used with great success on the MSSTA Ritchey- Chrétien telescope systems in assuring the accuracy of the systems prior to and supportive of other performance tests. It is also of extreme importance to employ these tests during the manufacturing of both the individual components as well as the final system. The use of multi-pass interferometry was made easier due to the highly reflective surface of the multi-layers even at the

longer test wavelength. This has made analysis more exact and the assessment of the performance at the shorter wavelengths has become more reliable.

## APPLICATIONS TO CANCER RESEARCH

Our current knowledge of tumor cell biology, detection and diagnosis has been made possible by steady improvements in microscopic methods for examining cells and tissues. Knowledge of cell ultrastructure made possible by high resolution transmission electron microscopy has revolutionized our concept of the organization of eukaryotic cells and the identification of organelles in the nucleus and cytoplasm. Recently, improvements in visible light instrumentation has led to the development of enhanced methods for the study of tumor cell growth and malignancy, and pre-malignant changes. For example, fluorescence microscopy, phase microscopy, differential interference contrast and polarizing microscopy have provided new approaches for research into pre-malignant changes and improvements in diagnosis. Moreover, the availability of low light level video cameras and computer enhancements of digitized images has led to improved resolution of the structures in motile cells. All current optical techniques are limited due to relatively low resolution and contrast of biological materials and the need, in most cases, for harsh fixatives, dyes and chemical additives.

The Water Window Imaging X-Ray Microscope has several potential features which could revolutionize tumor cell biology and cancer diagnosis. Its unique potential for detecting structures with spatial resolution in the object plane of 100 Å or better, along with its capacity to image living cells in aqueous, physiological environments is an advantage which is not available in any conventional microscopes. Although initial trials would be limited to cell monolayers in tissue culture, imaging could be extended to living tumor cells as well as analysis of frozen sections. In conjunction with tumor and cellular biology specialists, cell samples will be mounted and images obtained at x-ray wavelengths in the water window portion of the spectrum using high intensity synchrotron radiation sources. Computer analysis of the images will be carried out. These detailed laboratory/clinical tests and evaluation studies will provide the necessary data to establish the applicability of the Water Window Imaging X-Ray Microscope to cancer research.

## POTENTIAL FOR COMMERCIALIZATION

The successful completion of this project will result in the production of an entirely new imaging x-ray instrument which should permit the study of living tumor cells with unprecedented spatial resolution. It will allow the structures within the cells to be studied without the introduction of fluorescent dyes or chemical additives, which may alter the cellular processes under study. If the full optical potential of the microscope is realized, the *Water Window Imaging X-Ray Microscope* could become a standard diagnostic instrument in hospitals and medical centers throughout the world. It is envisioned that commercial instruments could be integrated with compact, internally contained, high intensity laser plasma x-ray sources. Its potential as an essential microscope for fundamental cell research would insure its commercial application in research centers and universities. Additionally the successful completion of this project will provide the impetus for the development of other complex advanced multilayer x-ray optical systems, of great importance to a broad range of other areas of science and technology.

This work is supported by MSFC Center Director's Discretionary Fund and NASA Grant NSG-5131, and the U.S. Dept. of Energy through Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

## REFERENCES

1. Walker, A. B. C., Jr.; Barbee, T. W., Jr.; Hoover, R. B.; and Lindblom, J. F.: Soft X-Ray Images of the Solar Corona with a Normal-Incidence Cassegrain Multilayer Telescope. *Science* vol 241, 1988, pp. 1781-1786.
2. Lindblom, J. F.; Walker, A. B. C. Jr.; Hoover, R. B.; Barbee, T. W., Jr.; VanPatten, R. A.; and Gill, J. G.: Soft X-Ray/Extreme Ultraviolet Images of the Solar Atmosphere with Normal Incidence Multilayer Optics. *Proc. SPIE*, vol 982, 1988, pp. 316-324.
3. Barbee, T. W., Jr.: Sputtered Layered Synthetic Microstructure [LSM] Dispersion Elements. AIP Conference Proc. No. 75, AIP, New York (D. Attwood and B. L. Henke, eds.), 1981. pp. 131-143.
4. Spiller, E.: Evaporated Multilayer Dispersion Elements for Soft X-Rays. AIP Conference Proc. No. 75, AIP, New York, (D. Attwood and B. L. Henke, eds.) 1981, pp. 124-131.
5. Walker, A. B. C., Jr.; Lindblom, J. F.; O'Neal, R. H.; Allen, M. J.; Barbee, Jr., T. W.; and Hoover, R. B.: Stanford/MSFC Multi-Spectral Solar Telescope Array. *Optical Engineering*, vol 29, 1990, pp. 581-591.
6. Hoover, R. B.; Baker, P. C.; Hadaway, J. B.; Johnson, R. B.; Peterson, C.; Gabardi, D. R.; Walker, A. B. C., Jr.; Lindblom, J. F.; and deForest, C.: Performance of the Multi-Spectral Solar Telescope Array III: Optical Characteristics of the Ritchey-Chrétien and Cassegrain Telescopes. *Proc. SPIE* vol 1343, (Richard B. Hoover and Arthur B. C. Walker, Jr., eds.) 1990, in press.
7. Baker, P. C.: Advanced Flow Polishing of Exotic Optical Materials. *Proc. SPIE*, vol 1160 (Richard B. Hoover, ed.), 1989, pp. 263-270.
8. Barbee, T. W., Jr.: Multilayers for X-Ray Optics. *Optical Engineering*, vol 25, 1986, pp. 898-915.
9. Rosenbluth, A. E.: Computer Search for Layer Materials that Maximize the Reflectivity of X-Ray Multilayers. *Revue Phys. Appl.* vol 23, 1988, pp. 1599-1621.
10. London, R. A.; Rosen, M. D.; and Trebes, J. E.: Wavelength Choice for Soft X-Ray Laser Holography of Biological Samples. *Appl. Opt.* vol. 28, 1989, pp. 3397-3404.
11. Spiller, E.: A Scanning Soft X-Ray Microscope Using Normal Incidence Mirrors. In *X-Ray Microscopy* (G. Schmahl and D. Rudolph, eds.), Springer-Verlag (Berlin), 1984, pp. 226-231.
12. Trail, J. A.; and Byer R. L.: X-Ray Microscope Using Multilayer Optics with a Laser Induced Plasma Source. *Proc. SPIE*, vol 563, 1985.
13. Lovas, I.; Santy, W.; Spiller, E.; Tibbits, R.; and Wilczynski, J.: Design and Assembly of a High Resolution Schwarzschild X-Ray Microscope for Soft X-Rays. *Proc. SPIE*, vol 316 (E. Spiller, ed.) 1982, pp. 90-97.
14. Hoover, R. B.; Shealy, D. L.; Gabardi, D. R.; Walker, A. B. C., Jr.; Lindblom, J. F.; and Barbee, T. W. Jr.: Design of an Imaging Microscope for Soft X-Ray Applications. *Proc. SPIE* vol 984 (F. E. Christensen, ed.), 1988, pp. 234-246.
15. Shealy, D. L.; Gabardi, D. R.; Hoover, R. B.; Walker, A. B. C., Jr.; Lindblom, J. F.; and Barbee, T. W., Jr.: Design of a Normal Incidence Multilayer Imaging X-Ray Microscope. *Journal of X-Ray Science and Technology*, vol 1, 1989, pp. 190-205.
16. Hoover, R. B.; deForest, C.; Lindblom, J. F.; O'Neal, R. H.; Peterson, C.; Walker, A. B. C., Jr.; and DeWan, A.: Performance of the Multi-Spectral Solar Telescope Array VI: Performance of the Photographic Film. *Proc. SPIE*, vol 1343 1990, in press.
17. Shealy, D. L.; Jiang W.; and Hoover, R. B.: Design and Analysis of Aspherical Multilayer Imaging X-Ray Microscope. *Proc. SPIE*, vol 1343 (Richard B. Hoover and Arthur B. C. Walker, Jr., eds.), 1990, in press.

## MECHANICAL RESPONSE TISSUE ANALYZER FOR ESTIMATING BONE STRENGTH

Dr. Sara B. Arnaud  
Research Scientist  
Ames Research Center

Dr. Charles Steele  
Professor of Applied Mechanics  
Stanford University

Anthony Mauriello  
President  
GaitScan, Inc.

## INTRODUCTION

One of the major concerns for extended space flight is weakness of the long bones of the legs, composed primarily of cortical bone, that functions to provide mechanical support. The strength of cortical bone is due to its complex structure, described simplistically as cylinders of parallel osteons composed of layers of mineralized collagen. The reduced mechanical stresses during space flight or immobilization of bone on Earth reduces the mineral content, and changes the components of its matrix and structure so that its strength is reduced (1). Currently, the established clinical measures of bone strength are indirect. They are based on determinations of mineral density by means of radiography, photon absorptiometry, and quantitative computer tomography. While the mineral content of bone is essential to its strength, there is growing awareness of the limitations of the measurement as the sole predictor of fracture risk in metabolic bone diseases, especially osteoporosis (2).

Other experimental methods in clinical trials that more directly evaluate the physical properties of bone, and do not require exposure to radiation, include ultrasound, acoustic emission, and low-frequency mechanical vibration. The last method can be considered a direct measure of the functional capacity of a long bone since it quantifies the mechanical response to a stimulus delivered directly to the bone. A low frequency vibration induces a response (impedance) curve with a minimum at the resonant frequency, that a few investigators use for evaluation of the bone (3). An alternative approach, the method under consideration here, is to use the response curve as the basis for determination of the bone bending stiffness  $EI$  and mass, fundamental mechanical properties of bone (4).

## DEVELOPMENT OF THE METHOD

Work was initiated by Young and Thompson at Ames Research Center to evaluate the effect of experimental disuse osteoporosis in a non-human primate model for weightlessness (5). The device they developed was an electromagnetic shaker that is placed on the middle of the bone to be tested. They selected the ulna in the arm and the tibia in the leg since both bones have central areas fairly close to the skin surface without a mass of muscle tissue between bone and the probe. A firm clamp held the limb to be tested in a sedated animal, as illustrated in Figure 1. The impedance head is attached to the shaker and the probe placed on the skin is attached to the impedance head so that the stimulus and measurement of the force and acceleration occur at one point.

## STUDIES IN THE NON-HUMAN PRIMATE

The force of the vibration causes a deflection from which the average cross-sectional bending stiffness,  $EI$ , can be calculated.  $E$  is the intrinsic material property and  $I$  is the cross-sectional moment of inertia. From this and the length of the bone, the limiting buckling force,  $P_{cr}$  can be computed ( $P_{cr} = EI (\pi/L)^2$ ). Bending stiffness in Kilo newtons per meter, measured in vivo by this impedance probe technique in the ulna and tibia

of the immobilized and recovering monkey are illustrated in Figure 2a. The pattern of decrease in bending stiffness in the left (closed circles) and right (open circles) tibias of 2 adult male monkeys in the absence of weight bearing was consistent, however, the ulna (triangles) in the upper extremity was not. The deficit in strength or bending stiffness present after 6 months required a longer period of time for restoration, i.e., 8 months.

Comparison of the loss of strength in the tibias during disuse seemed to parallel the decreases in bone mineral content, measured by single photon densitometry (Norland) and illustrated in Figure 2b. However, during the 12 month recovery period, there is a clear lag in restoration of mineralization. Although stiffness in the tibias of the animal in the lower panel returned to pre-disuse levels, mineral content remained low for 12 months. This preliminary data emphasizes the importance of direct measures of strength as well as mineral content.

### ADVANCES IN THE ANALYSIS SYSTEM

These animal studies coincided with advances in the analysis system to enable a more rapid and efficient measurement. A microprocessor system for analyzing the impedance curves, the "Steele Oxbridge Bone Stiffness Analyzer" was developed to compute the data in seconds instead of hours. The system used a dual channel dynamic signal analyzer (HP3562A) and computer (HP9826A) with dual 3-1/2 inch disk drives and equipment from Bruell and Kjaer for the impedance head (BK4810 - permanent magnet vibration exciter, BK8001 - impedance head, and BK2635 - charge amplifiers). Instead of the standard impedance curve, Steele and his colleagues found it better to work with the ratio of force to displacement obtained from impedance by multiplying by frequency, in the range 0-1600Hz. The result is then expressed in Newtons per meter stiffness or force/displacement.

The resolution of responses from bone was complicated by the need to subtract the resonance from skin and to recognize the damping effect of muscle and resonances from the ends of the bone when they are not firmly positioned. These hurdles were overcome sufficiently to determine the stiffness of the human ulna and verify the theoretical concepts.

### BENDING STIFFNESS IN THE HUMAN ULNA

Sufficient data from 4 testing sites have generated results for the ulna that establish the interassay variation between 4% and 5%, an acceptable range for testing on the same day. Both the ability of the test subject to relax and the experience of the operator in positioning the arm influence this considerably. With either a hand held probe or a fixed probe, the variation in the measurement over a month's time averages about 8 percent. It is still uncertain as to whether this variability in subject repeatability is due to actual physiologic changes in the bone or to differences in positioning. Work to automate positioning to reduce this interassay variation from month to month is in progress.

Improved repeatability of the test will be essential for the evaluation of the changes in bone strength that occur in subjects during a ground-based model for weightlessness. Preliminary data in 11 right-handed volunteers for a bed rest study at Ames Research Center indicated biologically valid changes in the stiffness of the ulna. There were decreases in the stiffness of the right ulna in 9 of 11, and increases in stiffness in the left ulna in 8 of 11. These changes were entirely consistent with our current concepts of redistribution of bone mineral in this disuse model and losses in areas of the skeleton where normal function has been disrupted.

The most compelling data for justifying continued efforts to pursue the development of the MRTA has come from the extensive testing of a research instrument in healthy adults and patients at Stanford University and the Palo Alto Veterans Administration Hospital. The relationship of stiffness of the ulna to the bone density of the radius, the adjacent and parallel arm bone, show good correlations (approximately 0.8) that confirm our knowledge of the relationship of strength of bone to the mineral content, we assume to be similar in the radius and ulna. The most important observation for future application of the test has come from a comparison of the axial load capability of the ulna, Pcr, determined from the lateral stiffness measurement,



and the bone mineral content of physically active and inactive subjects, as shown in Figure 3. The load capability is higher in the active than inactive subjects for the same value of bone mineral. This data emphasizes the importance of the quality of bone for its functional capability, as well as its mineral content.

### CLINICAL INSTRUMENTATION

The most recent advances in the development of the MRTA have been in the design and construction of a clinically useful instrument carried out by Gaitscan, Inc. Automatic centering of the probe on the long bone to be tested and an adjustable limb support that can be easily raised, lowered or moved horizontally to accommodate the variety in body types, has greatly reduced the time of the test and increased subject acceptance. The design of the limb support has enabled us to acquire data in the tibia, the bone in the leg that is most vulnerable to space flight. The more irregular shape of the tibia than the ulna or other factors make it a more difficult bone to measure with reliability. Nevertheless, our preliminary data shows a good correlation of stiffness with bone mineral content and stiffness values in the range theoretically estimated for a beam of larger dimensions and mass than the ulna. Refinements of instrument design and of the data analysis to reduce the variation in the test results, higher in the tibia than in the ulna, are in progress. With this newly constructed instrument it is possible to test large numbers of people.

### APPLICATIONS OF BONE STIFFNESS TESTING

The application of a non-invasive direct measurement of the strength of long bones is most obviously to the identification of individuals with deficits due to a variety of causes, and to the monitoring of treatment programs aimed at correcting the deficits. While there has been considerable interest in the diagnosis and prediction of fracture in patients with senile or post-menopausal osteoporosis, one of the major national health problems, there is now no firm data that indicates the value of this measure of bone strength to be superior to the newer radiographic methods for diagnosis. Bone diseases in which trabecular bone is the primary target would require extensive clinical testing in patients with evidence of cortical bone involvement to evaluate the utility of the MRTA. The more immediate applications of the MRTA in medicine are in the follow-up of osteoporotic patients or others receiving various treatments directed at improving deficits in mineral and/or strength as well as patient activity. The effects of fitness programs carried out for any purpose at any age and patient population can be easily monitored with the MRTA.

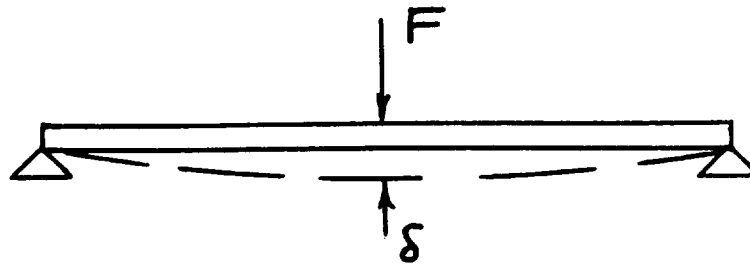
Aside from the importance of a direct measure of a physical property of bone to the astronaut, the safety and simplicity of the bone strength test lends itself to screening studies aimed at identifying substandard individuals, with respect to bone quality. Surveys of military recruits, high school and college age students would be particularly informative for those interested in the musculoskeletal fitness status of the nation.

### REFERENCES

1. Patterson-Buckendahl, P., Arnaud, S. B., Mechanic, G. L., Martin, R. B., Grindeland, R. E., and Cann, C. E. Fragility and composition of growing rat bone after one week in spaceflight. *Am. J. Physiol.* 252 (Regulatory Integrative Comp. Physiol. 21): R240-R246, 1987.
2. Ott, S. M., Kilcoyne, R. F., and Chestnut, C. H. III. Comparisons among methods of measuring bone mass and relationship to severity of vertebral fractures in osteoporosis. *J. Clin. Endocrinol. Metab.* 66: 501-507, 1988.
3. Doherty, W. P., Bovill, E. G., and Wilson, E. L. Evaluation of the use of resonant frequencies to characterize physical properties of human long bones. *J. Biomechanics* 7: 559-561, 1974.
4. Steele, C. R., Zhou, L.-J., Guido, D., Marcus, R., Heinrichs, W. L., and Cheema, C. Noninvasive determination of ulnar stiffness from mechanical response -- *in vivo* comparison of stiffness and bone mineral content in humans. *ASME J. Biomech. Eng.* 110: 87-96, 1988.

5. Thompson, G. A., Orne, D., and Young, D. R. In vivo determination of mechanical properties of the human ulna by means of mechanical impedance tests: Experimental results and improved mathematical model. *Med. Biol. Eng.* 14: 253-262, 1976.
6. Young, D. R., Howard, W. H., Cann, C., and Steele, C. R. Noninvasive measures of bone bending in the monkey (*M. nemestrina*). *Calcif. Tissue* 27: 109-115, 1979.
7. Young, D. R., Niklowitz, W. J., and Steele, C. R. Tibial changes in experimental disuse osteoporosis in the monkey. *Calcif. Tissue Int.* 35: 304-308, 1983.
8. Steele, C. R. and Gordon, A. F. Preliminary clinical results for "SOBSA" noninvasive measurement of ulnar bending stiffness. *ASME Advances in Bioengineering*, R. C. Eberhard and A. H. Burstein, eds., pp 85-87.
9. Arnaud, S. B. and Morey-Holton, E. Gravity, Calcium and Bone: Update, 1989. *The Physiologist* 33: S65-S68, 1990.

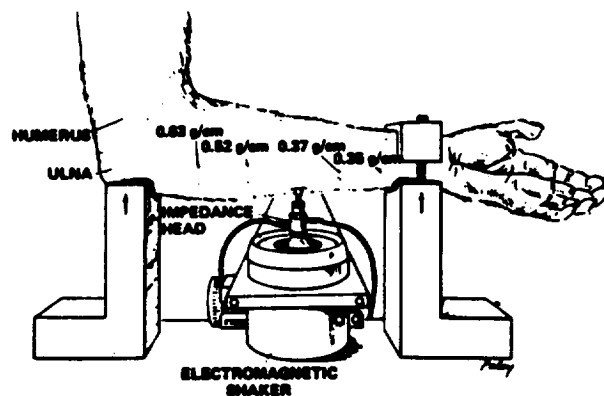
**FIGURE 1.** From Steele, C. R., Proc. Workshop on Advances in NASA-Relevant Minimally Invasive Instrumentation, 1984. JPL D-1942.



A beam, supported on the ends, when loaded by the force  $F$  will deflect the distance  $\delta$ . The lateral stiffness is  $k = F / \delta$ .



A beam with compression force  $P$ . The limiting force is the buckling load  $P_{cr}$ , which can be computed from the lateral stiffness.



Noninvasive measurement of ulnar stiffness in monkey.  
(from reference 6)

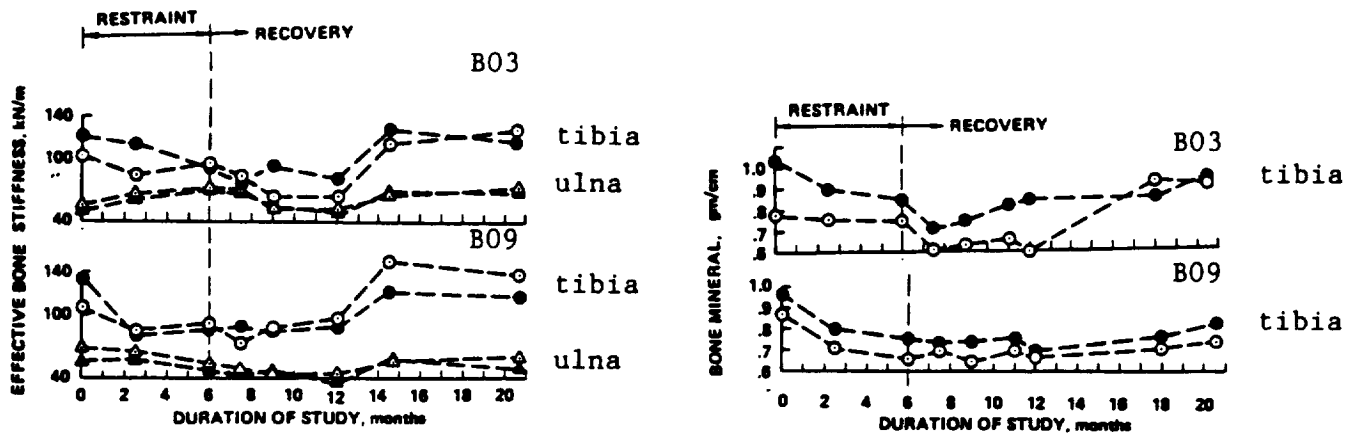
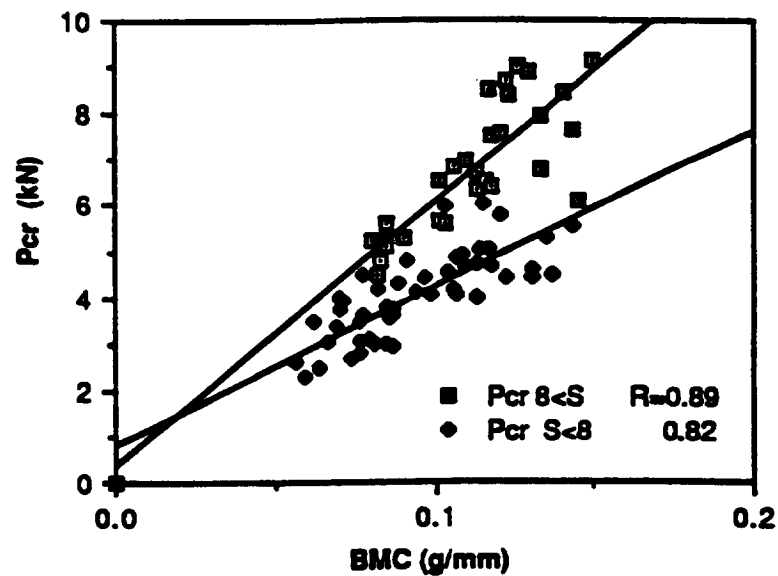


FIGURE 2. Adapted from reference 6



Comparison of axial load capability of ulna Pcr, determined from lateral stiffness measurement, and bone mineral content.

FIGURE 3. From reference 6

## ADAPTATION OF NASA TECHNOLOGY FOR THE OPTIMIZATION OF ORTHOPEDIC KNEE IMPLANTS

by

D.A. Saravanos<sup>1</sup>, P.J. Mraz<sup>1</sup>, D.T. Davy<sup>1</sup>, and D.A. Hopkins<sup>2</sup>

National Aeronautics and Space Administration  
Lewis Research Center  
Cleveland, OHIO 44135

### ABSTRACT

NASA technology originally developed for the optimization of composite structures (engine blades) is adapted and applied to the optimization of orthopedic knee implants. A method is developed enabling the tailoring of the implant for optimal interaction with the environment of the tibia. The shape of the implant components are optimized, such that the stresses in the bone are favorably controlled to minimize bone degradation and prevent failures. A pilot tailoring system is developed and the feasibility of the concept is evaluated. The optimization system is expected to provide the means for improving knee prosthesis and individual implant tailoring for each patient.

### INTRODUCTION

The history of design of orthopedic implants represents an extended iterative process of design improvements. Particularly for total hip replacement (THR) and total knee replacement (TKR) implants (Fig. 1), designs have improved to the point where probabilities for good long-term survival rates and functional restoration are quite high. Nonetheless, higher long term success rates and longer potential useful lives are clearly desirable goals.

The problem of optimizing joint prosthesis design is most commonly defined in terms of structural mechanics. Although unacceptable clinical results can occur without structural failure, the predominant modes of implant failure may be described in terms of structural mechanics. Huiskes [1] has noted five possible design optimization objectives, all of which relate to possible structural failures in either the interface between component and adjacent bone, the adjacent bone itself, or the implant component.

Given that the prosthetic joint will satisfactorily reproduce gross joint dynamics, the goal of design optimization becomes one of optimizing the shape and the material characteristics of the implant components to control stresses or other related quantities in the structure. The satisfactory definition of the optimization criteria can be a formidable task. However, once this is accomplished, the remainder of the design optimization process then becomes the laborious process of searching for the best corresponding choices for implant geometry and materials. The

---

<sup>1</sup> Resident Research Associate, Graduate Research Assistant, and Professor of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OHIO.

<sup>2</sup> Aerospace Engineer, NASA Lewis Research Center, Cleveland, OHIO.

use of numerical nonlinear programming techniques for this purpose has proven to be an extremely useful tool to the designer. Coupled with advanced structural analysis techniques, especially the finite element method (FEM), they have been used extensively with considerable success in structural design [2].

Among the many contributions in the field, we selectively mention some of the work performed at NASA-Lewis on the multi-objective and multi-disciplinary optimization of layered composite structures [3, 4], as it is related to the research described herein. Applications of analogous optimal design methodologies in biomechanical design have been relatively few and limited in scope [1, 5-7]. In the present paper, we report on some recent work which has been conducted by applying the previously mentioned composite mechanics and structural optimization techniques to the design of the tibial component of TKR. The long-term goals are to develop some broadly applicable methodologies for optimizing the design of orthopedic implants. The first efforts have been concentrated on examining the capabilities and limitations of these methods in application to shape optimization procedures in optimizing total knee prosthetic component shapes.

The methodology and the computer code are adapted from those used for the multi-disciplinary structural tailoring of composite structures [3, 4]. While there are many obvious differences in the applications, the combination of essential features, namely the composite mechanics, the FEM models, the stress analysis, and the structural optimization techniques remains the same.

As a starting place, a pilot tailoring system is developed for the optimal design of the tibial component of the TKR. Specifically, we have considered the metal backed component with a central post or stem, and have examined the shape tailoring of the post and the thickness sizing of the metal backing or tray. In our approach, we assume, as have others [1, 5-7], that the basic criterion for design optimization is the stress state in the bone/implant composite structure. For the tibial component, the most likely mode of failure is aseptic loosening, which involves the resorption of the bone adjacent to the implant. This is presumably due in many cases to the stress-induced adaptation. Thus, the optimal design goal becomes the minimization of undesirable bone atrophy of the adjacent bone. In this paper, we report the basic features of the methodology and its evaluation of the shape tailoring of a tibial prosthesis using this approach.

## STRUCTURAL ANALYSIS

The present section outlines the NASA technology which was used to model the quasi-static response of the upper tibia. The technology has been successfully applied in the analysis of composite laminates and structures [3, 8], but has seen limited application in the case of bone structures despite the profound similarities, namely, the capability to represent nonuniform material regions and out-of-plane bending/torsional deformations. Instead, most of the efforts have been focused either on two-plane finite element analysis based on thickness corrections to represent the nonhomogeneity of the implanted tibia or on more elaborate three-dimensional analyses of the continuum solid. The proposed laminate theory may be viewed as a compromise between these two extremes.

The analysis involves three stages. In the first stage (upward synthesis), the stiffness characteristics of the discrete bone layers are synthesized. Integration through the thickness (sagittal plane) provides the equivalent extensional and flexural stiffness properties, relating the average (generalized) laminate stresses to the average (generalized) laminate strains. In the second stage, the global static response is obtained with finite element analysis. The third and final stage involves the back-calculation of stress and strain at each layer from the global structural response.

## OPTIMAL IMPLANT DESIGN

The objective of the proposed design procedure is to control the stress field in the cancellous and cortical bone of the implanted tibia by changing the shape and the dimensions of critical implant components. The original (unimplanted) tibia has a unique bone macro- and micro-morphology to efficiently carry the applied loads at the joint. It is generally accepted [9-13] that this particular bone morphology has been developed in response to an average stress/strain stimulus induced by the applied joint loads. The bone adapts its morphology reacting to this stress stimulus. The presence of the prosthesis alters both the mechanical characteristics and the load path in the implanted tibia, hence, it significantly perturbs the stress field creating the potential for changes in the bone, which may cause bone atrophy in the vicinity of the bone-implant boundaries. The latter phenomenon is undesirable in knee prosthesis since it is primarily responsible for the typical failure mode in this joint, that is, the loosening and subsequent failure of the tibial component.

Based on experimental results and clinical observations, phenomenological models have been developed to the point where one can attempt numerical predictions of bone remodeling histories [9, 10]. However, considerable work remains to validate the various models, and the appropriate parameters to be incorporated into these models are still to be determined experimentally. Hence, optimizing prosthesis design using time dependent remodeling is relatively impractical at the present time. The alternative approach, using a stress- or strain-related criterion, has a direct relationship to the anticipated remodeling behavior since they are typically the basis for the remodeling models [11-13]. The use of a stress-based criterion, although it does not follow the history of the remodeling process nor does it reflect the time dependent adaptation, is nevertheless sufficient for a first approach to the design of an optimal prosthetic implant.

It appears that the most important requirements are minimal interfacial movement and minimal bone stresses/strains near the bone/implant interfaces [12, 14]. The global stress field in the implanted tibia should be also maintained within normal physiological levels. In this way, global alterations in the bone tissue will be retained in minimal levels.

Other secondary modes of implant failures may be either bone fracture or mechanical failure of the implant components as a result of high stress concentrations in extreme loading conditions. Although preliminary FE analyses indicate that the stress level in the bone and the implant are too low for this mode of failure, suitable preventive mechanisms have been incorporated into the proposed methodology.

In view of the previous observations, candidate quantitative design criteria leading to favorable stress fields within the implanted tibia are: (1) minimization of the stress concentrations at the vicinity of the bone/implant interface, and (2) maintenance of the stress in the bone within the physiological levels of the original tibia. Some additional criteria involve the prevention of mechanical failures, bounds on the admissible shape for the tibial post, and minimal removal of bone tissue.

This constrained optimization problem was solved numerically with the modified feasible directions non-linear programming method [15]. The feasible directions method is a direct search algorithm performing direct search in the feasible design domain. The search direction is estimated from the gradients of the objective function and the active constraints. A line search along this search direction based on polynomial interpolation is then performed to complete each design move.

## **APPLICATION AND RESULTS**

The present section includes evaluations demonstrating the applicability of the tailoring system on the shape optimization of a single-post tibial component (Fig. 1) subject to symmetric and unsymmetric loading in the frontal plane. The section also describes the geometry and material properties and the validation of the FEA model.

### **Geometry and Material Properties**

Typical dimensions and bone morphology representing the upper tibia of an average adult male were assumed in this study. For simplicity, symmetry was assumed in both frontal and sagittal planes, as well as, elliptical cross-sections. Nevertheless, the developed pilot tailoring system entails the capacity to handle more complicated bone geometries and morphologies. The variation in bone material properties was approximated by discrete laminated material regions and properties as shown in Fig. 2. As seen in Fig. 2, the discretized bone morphology consists of cortical bone surrounding different densities of cancellous bone.

The tibial component of the prosthetic knee implant, which was used both as a reference and an initial design, is a simple but valid representation of ones found to be in common use today and therefore it is a suitable start for the optimization process. The reference design incorporates a 15 mm thick ultra-high molecular-weight polyethylene (UHMWPE) plateau supported by a 3 mm thick titanium backing tray. A 15 by 15 mm square by 30 mm long titanium stem or post protrudes from the bottom center of the tray. The polyethylene plateau and the titanium backing tray are both 70 mm wide and 45 mm deep and they match exactly the geometry of the corresponding tibial cross-section.

### **Loading Conditions**

Two different loading conditions were considered in this paper (Fig. 3):

- (1) Symmetric, in-plane, vertical forces parabolically distributed over each condylar surface area totaling 2000 N. The load magnitude of 2000 N represents a resultant joint reaction force equivalent to three times the body weight, which is typical during normal, level walking [16, 17].
- (2) Unsymmetric, in-plane, vertical forces totalling 1333 N and 667 N (2:1 ratio) respectively, parabolically distributed over their corresponding condylar surface. Unsymmetric loads are considered to be a more realistic representation of the actual loading conditions within the knee joint.

### **Validation of the FEA Model**

The validity and limitations of the finite element model incorporated in the tailoring system were illustrated by comparisons with more detailed three-dimensional continuum finite element models of the same implanted tibia. The comparisons made are for the reference prosthesis design and involved the same symmetric and unsymmetric in-plane loading conditions as described earlier. There is good agreement between the two models, which is evident when comparing the stresses in each model for the cortical bone and cancellous bone regions as well as the tray and post configuration. While the stresses at every element are not exactly identical, the trends of the stresses in each model are comparable. That is, the general behavior of the stress contours in the three-dimensional model are adequately reproduced in the two-dimensional case for both symmetric and unsymmetric loading conditions.



## Results

Corresponding to the requirement for maintenance of the bone/implant interface, we have chosen the objective function to be minimization of the maximum stress at the bone/implant interface regions.

The design variable sets for the optimization cases presented here involved combinations of post and tray geometry parameters including: the post length (-15% to +20%), width tapering along the length of the post (-25% to +25%), and tray thickness (-3% to +10%). The values in parenthesis indicate the lower and upper bounds imposed as a percent of the respective reference implant design value.

**In-Plane Loading, Symmetric.** Shape optimization of the post only resulted in a 28% reduction in the maximum stress which occurs at the outermost elements in the tray/bone interface. For this design case, the post width has been tapered and its length increased.

Combined optimization of the post and tray resulted in a 30% reduction in the maximum stress and results in nearly the same stress distribution as in Case 1. The tray was thickened considerably illustrating that thicker trays provide improved immobilization at the tray/bone interface. The current trend of using metallic backing trays as opposed to the less successful all-plastic tibial implant configuration reinforces the obtained results. The combined optimization did not produce significant additional reductions in the maximum stress.

Thus, the results for optimization in a symmetric loading scenario indicate the need for a design that includes a thicker holding tray with a wider but tapered post of slightly longer length as compared to the initial design case. Such a design transfers the maximum stress from the outer portions of the tray/bone interface region to regions below the bottom of the post within the cortical shell (Fig. 4).

**In-Plane Loading, Unsymmetric.** Optimization of the post resulted in a 12% decrease in the maximum stress within the bone/implant interface, and the maximum stress has now been transferred to elements in the bone which are below the lower end of the post. The post has been slightly shortened and widened as well as tapered.

Optimization of the post and the tray resulted in a 18% reduction in the maximum stress within the bone/implant interface. In this case, the post has been shortened and tapered while the tray has been made thicker.

Thus, the prosthesis optimization under unsymmetric loading indicates the need for a design that includes a thicker holding tray with a wider but tapered post of slightly shorter length as compared to the initial design case. Such a design results in the higher stresses appearing in regions that are below the bottom of the post (Fig. 5).

## SUMMARY

A method is developed for the optimization of orthopedic knee implants by adapting NASA technology originally developed for the analysis and optimization of composite structures. The present paper is focused on the optimal design of the tibial implant components. The shape of the implant is tailored for improved bone growth near the boundaries of the implant. A research pilot code was developed to demonstrate the feasibility of the concept.

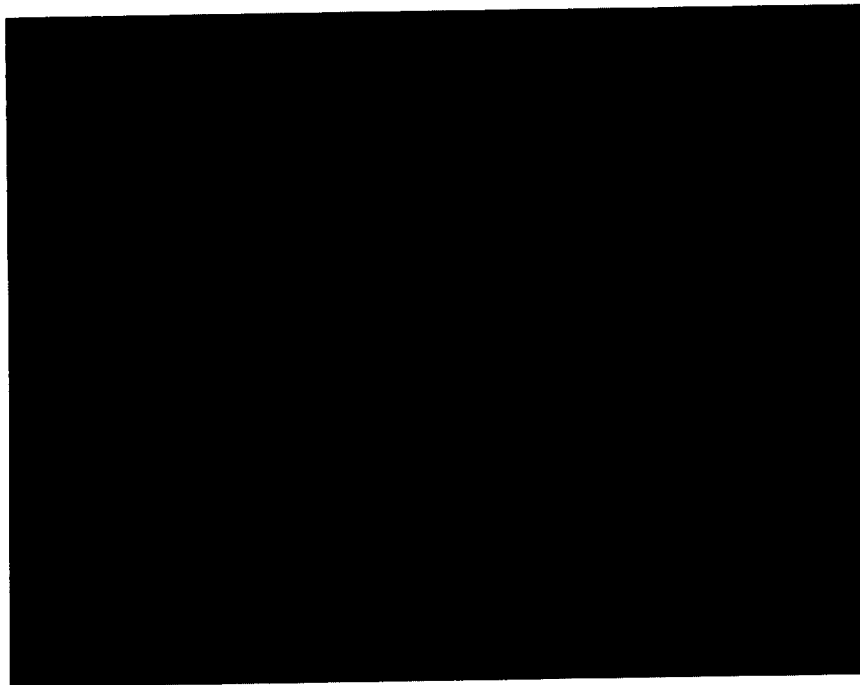
Evaluations of the pilot tailoring system were performed on the shape optimization of tibial implants subject to symmetric and unsymmetric distributed loads applied on the epiphyseal plate. The results illustrated that optimization of the implant shape can indeed reduce the maximum stress in the bone at the vicinity of the implant, therefore, shape optimization techniques can be successfully applied in designing durable and customized knee implants. The applications also demonstrated the dependence of the resultant optimal designs to the loads, suggesting the need for optimization under multiple loading conditions. The material properties of the implant were also proved important. Optimization of the material properties will be best accomplished with tailored composite implants, and work in this subject is currently in progress.

**Acknowledgements** -- The authors wish to express their thanks to Dr. Christos C. Chamis of the NASA Lewis Research Center for his direction and advisement. This work is supported by the technology utilization office at the NASA Lewis Research Center, Cleveland, OH under NASA Grant NAG 3-1027.

## REFERENCES

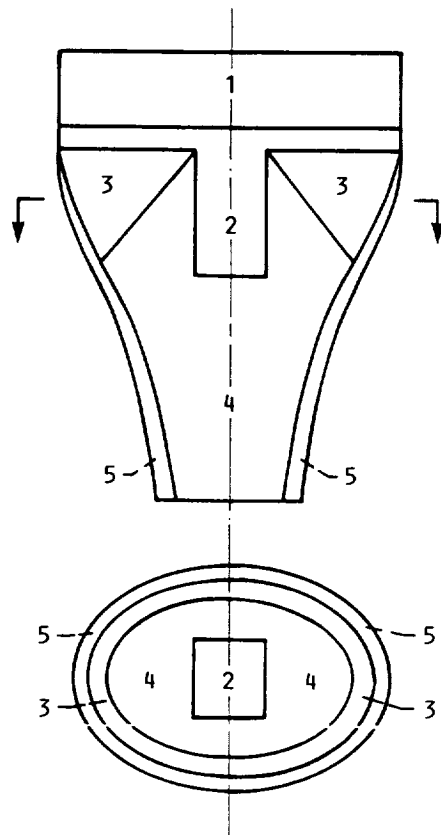
- 1 Huiskes, R.; Boeklagen, R.: Mathematical Shape Optimization of Hip Prosthesis Design. *J. Biomechanics*, vol 22, 1989, pp. 793-804.
- 2 Haftka, R.T.; Grandhi, R.V.: Structural Shape Optimization - A Survey. *Computer Methods in Applied Mechanics and Engineering*, vol 57, 1986, pp. 91-106.
- 3 Brown, K.; Harvey, P.: Statistical Tailoring of Advanced Turboprops (STAT). NAS3-23941, 1987.
- 4 Saravanos, D.A.; Chamis, C.C.: Multi-Objective Shape and Material Optimization of Composite Structures Including Damping. NASA TM 102579, 1990.
- 5 Huiskes R.; Boeklagen, R.: The Application of Numerical Shape Optimization to Artificial Joint Design. "Computational Methods in Bioengineering," (Spilker, R.L.; Simon, B.R., eds.) *BED*, vol 9, ASME, 1988, pp. 185-197.
- 6 Huiskes, R.; Kuiper, J.H.: Numerical Shape Optimization of Prosthetic Implants. *Proc. of the First World Congress of Biomechanics*, La Jolla, CA, 1990.
- 7 de Beus, A.M.; Hoeltzel, D.A.; Eftekhar, N.S.: Design Optimization of a Prosthesis Stem Reinforcing Shell in a Total Hip Arthroplasty. *J. biomech. Engng.*, vol 112, 1990, pp. 347-357.
- 8 Murthy, P.L.; Chamis, C.C.: Integrated Composite Analyzer (ICAN). NASA TP 2515, 1986.
- 9 Hart, R.T.; Davy, D.T.; Heiple, K.G.: A Computational Method for Stress Analysis of Adaptive Elastic Materials With a View Toward Applications in Strain-Induced Bone Remodeling. *J. of biomech. Engng.*, vol 106, 1984, pp. 342-350.
- 10 Beaupre, G.S.; Orr, T.E.; Carter, D.R.: An Approach for Time-Dependent Bone Modeling and Remodeling--Theoretical Development. *J. Orthopaedic Research*, vol 8, no 5, 1990, pp. 651-661.
- 11 Cowin, S.C.; Hart, R.T.; Balsar, J.R.; Kohn, D.H.: Functional Adaption in Long Bones: Establishing *in-vivo* values for surface remodeling rate coefficients. *J. Biomechanics*, vol 18, 1985, pp. 665-684.

- 12 Huiskes, R.; Weinans, H.; Grootenboer, H.J.; Dalstra, M.; Fudala, B.; Sloof, T.J.: Adaptive Bone-Remodeling Theory Applied to Prosthetic-Design Analysis. *J. Biomechanics*, vol 20, 1987, pp. 1135-1150.
- 13 Carter, D.R.; Fyhrie, D.P.; Whalen, R.T.: Trabecular Bone Density and Loading History: Regulation of Connective Tissue Biology by Mechanical Energy. *J. Biomechanics*, vol 20, 1987, pp. 785-794.
- 14 Vasu, R.; Carter, D.R.; Schurman, D.J.; Beaupre, G.S.: Epiphyseal-Based Designs for Tibial Plateau Components--I. Stress Analysis in the Frontal Plane. *J. Biomechanics*, vol 19, 1986, pp. 647-662.
- 15 Vanderplaats, G.N.: A Robust Feasible Directions Algorithm for Design Synthesis. *Proc. 24th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics, and Materials Conference*, Lake Tahoe, NV, 1983.
- 16 Morrison, J.B.: The Mechanics of the Knee Joint in Relation to Normal Walking. *J. Biomechanics*, vol 3, 1970, pp. 51-61.
- 17 Smidt, G.L.: Biomechanical Analysis of Knee Flexion and Extension. *J. Biomechanics*, vol 6, 1973, pp. 79-92.



ORIGINAL PAGE IS  
OF POOR QUALITY

**Fig. 1** Typical components of knee implants (from left to right): (a) Femoral component; (b) Patellar component; (c) Tibial component.



MATERIAL PROPERTIES FOR CORRESPONDING MATERIAL REGIONS

MATERIAL	DESCRIPTION	YOUNG'S, MPa	MODULUS, Mpsi	POISSON'S RATIO
1	POLYETHYLENE	500	0.08	0.40
2	TITANIUM	113 764	16.50	0.30
3	CANCELLOUS HD	400	0.06	0.20
4	CANCELLOUS LD	200	0.03	0.20
5	CORTICAL SHELL	15 169	2.20	0.32

Fig. 2 Morphology and material regions for typical adult male tibia with prosthesis.

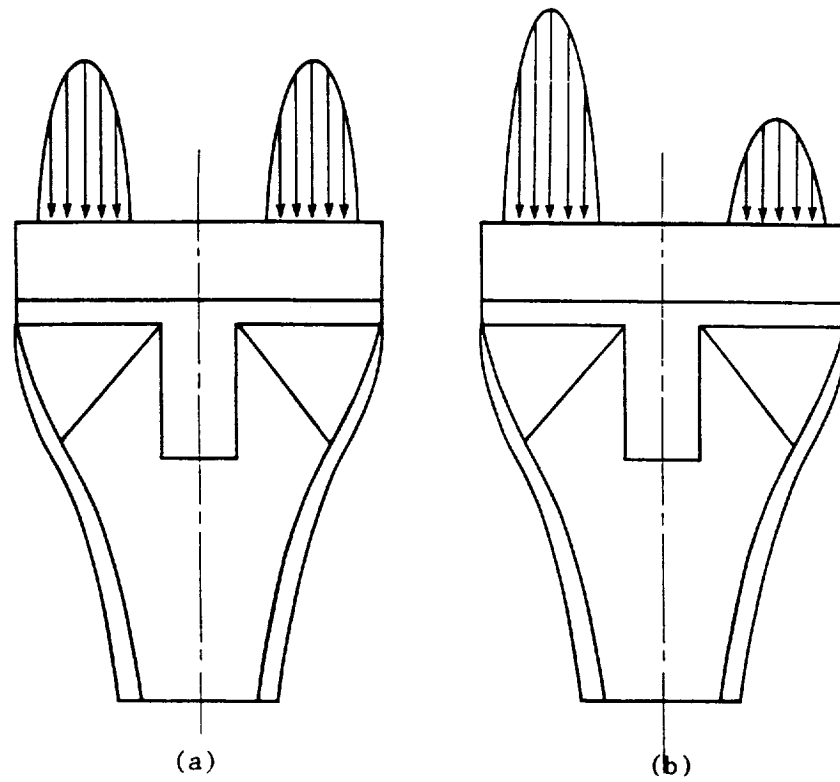
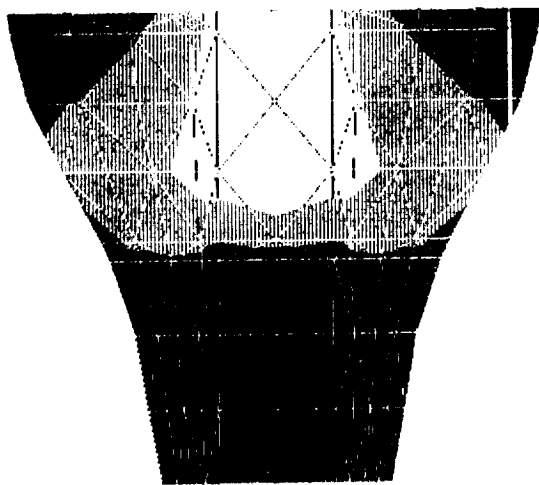
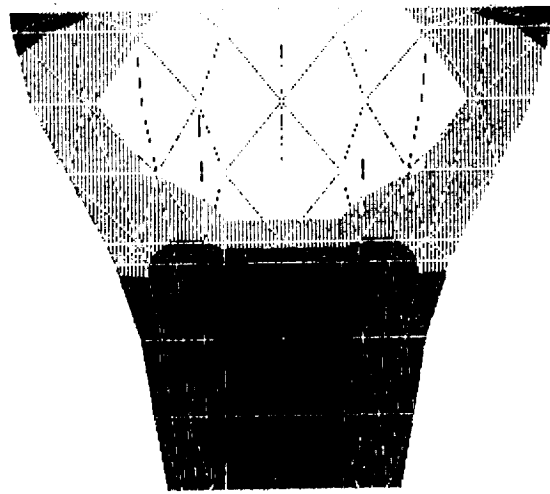


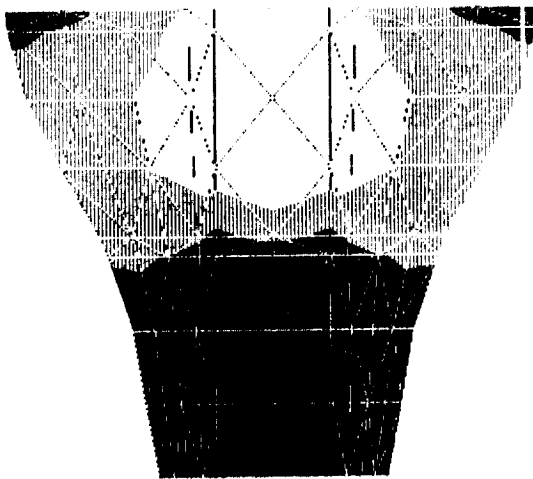
Fig. 3 Loading scenarios: (a) In-plane, symmetric; (b) In-plane, unsymmetric.



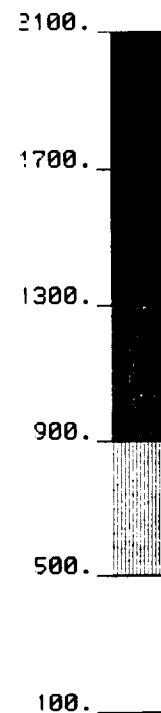
(a)



(b)

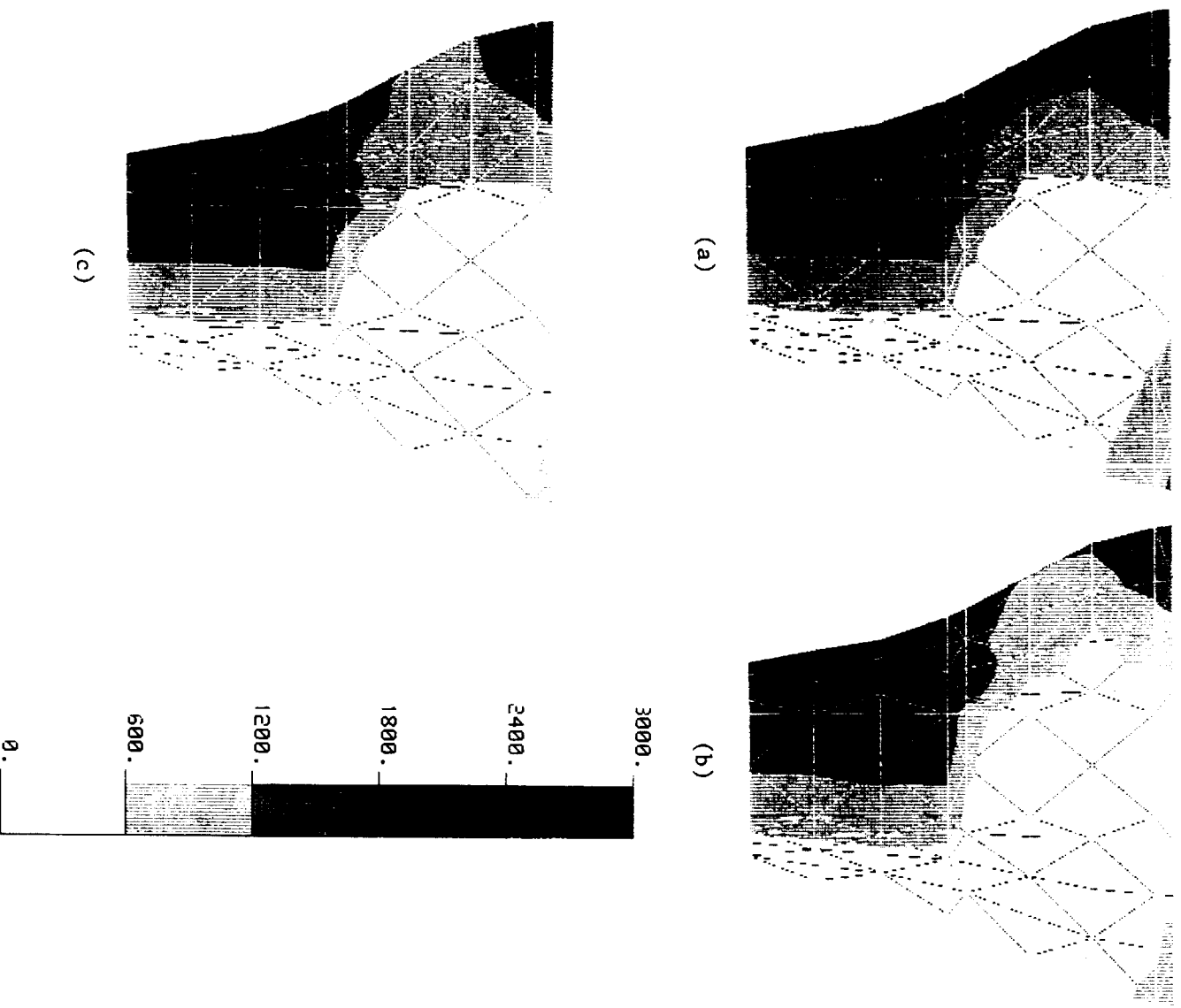


(c)



PRINTED PAGE 12  
OF 1000 QUALITY

**Fig. 4** Equivalent stress distributions in the cortical bone for symmetric loads: (a) Reference design; (b) Optimized post only; (c) Optimized tray and post.



**Fig. 5** Equivalent stress distributions in the cortical bone for unsymmetric loads:  
 (a) Reference design; (b) Optimized post only; (c) Optimized tray and post.

## **SESSION C - INFORMATION AND DATA MANAGEMENT**

**Tuesday November 27, 1990**

- **Optical Storage Device**
- **Rewriteable Optical Disk Recorder Development**
- **Monitoring And Analysis Of Data From Complex Systems**
- **High Data Rate Systems For The Future**
- **DAVID: The Distributed Access View Integrated Database**
- **Three-Dimensional Perspective Visualization**





## Optical Storage Device

Sharon S. Welch  
 Spacecraft Controls Branch  
 NASA Langley Research Center  
 Hampton, VA.

Summary

A new holographic image storage device which uses four-wave mixing in two photorefractive crystals is described. Photorefractive crystals promise information storage densities on the order of  $10^9$  to  $10^{12}$  bits per cubic centimeter at real-time rates. Several studies in recent years have investigated the use of photorefractive crystals for storing holographic image information. However, all of the previous studies have focused on techniques for storing information in a single crystal. The disadvantage of using a single crystal is that the read process is destructive. Researchers have developed techniques for fixing the information in a crystal so that it may be read many times. However, when fixed, the information cannot be readily erased and overwritten with new information. If two photorefractive crystals are used, holographic image information may be stored dynamically. That is, the stored image information may be read out more than once and, it may be easily erased and overwritten with new image information.

Introduction

In the mid 1960's, researchers discovered what has come to be known as the photorefractive effect. Certain types of crystals, when placed in the optical cavities of lasers, were found to suffer "optical damage" after a brief period of time (ref. 1). Later, it was determined that the crystals were in fact not damaged, but rather had the remarkable property that when exposed to light would undergo a change in refractive index. In the late 1960's, researchers realized that this property could be exploited and these crystals could be used to record holographic information (ref. 2). When exposed to multiple coherent interfering beams of light, these crystals will record (on a microscopic level) the interference patterns generated by the beams. These interference patterns (recorded as a change in the index of refraction) are phase holograms or holographic gratings.

The mechanism by which holograms are recorded in photorefractive crystals can be understood with a simple example. In figure 1 is a drawing showing two coherent plane wave beams of light with wave vectors  $\kappa_1$  and  $\kappa_2$  intersecting at some angle  $2\theta$  in a photorefractive crystal. The alternating dark and light bands within the region of intersection represent the light distribution resulting from interference of the two beams. The spacing of the bands is described by the spatial frequency  $k$ , where

$$k = 2 |\kappa_1| \sin \theta$$

For very small angles of  $\theta$ , the spacing of the bands of light will be very fine. On a microscopic level, the individual atoms of the crystal are exposed to alternating bands of light and dark. In the illuminated regions, light is absorbed, causing excitation of trapped charges within the crystalline structure. The excited charges migrate and become retrapped in regions of low light intensity (the dark bands). This results in charge separation within the crystal. This charge separation sets up a strong static electric field which in turn causes a change in the refractive index of the crystal due to the linear electro-optic effect (ref. 3). As shown in figure 2, the induced change in refractive index mimics the interference pattern except it is shifted in space  $\pi/2$  out of phase with respect to the incident light intensity pattern. The microscopic spatial variation in the refractive index of the crystal produced by the interference of two coherent beams of light is a type of hologram known as a phase hologram.

Holograms of objects differ from photographs in that they are not a record of the light intensity distribution reflected from an object, but rather a record of the amplitude and phase distribution of light

reflected from an object. When holograms are illuminated, the amplitude and phase of the wavefront of the light reflected from (or transmitted through) the original object are reconstructed. To an observer, this wavefront is indistinguishable from the wavefront reflected off the object itself and the observer therefore "sees" the original object in three dimensions. It is by illuminating the hologram that the information stored in the hologram is read out. Unlike a holographic grating recorded in film or other fixed media, when a holographic grating in a photorefractive crystal is illuminated, the charges in the crystal redistribute and the grating is erased. For this reason, the read process, if separated from the write process, is destructive.

The optical storage device described herein uses a four-wave mixing geometry to store holograms in two photorefractive crystals. In four-wave mixing, the read and write processes are combined. Three coherent beams of light are incident on a photorefractive crystal as shown schematically in figure 3. One beam, denoted as the object beam carries the information to be stored in the form of a complex phase and amplitude distribution. This beam is interfered with a plane wave reference, denoted as the write beam in figure 3, thereby generating a complicated interference pattern which is recorded in the crystal as described previously. A third beam, denoted as the read beam, is a plane wave reference beam counterpropagating with respect to the write beam. When the holographic grating is illuminated by the read beam, a portion of the read beam is diffracted by the grating and reconstructs the phase conjugate of the object beam. The phase conjugate is counterpropagating with respect to the object beam. It is a time-reversed replica of the object beam.

In the optical storage device, holographic information is recorded in two photorefractive crystals. An object wave bearing the information to be stored is interfered with a plane wave reference or write beam, and the resulting interference pattern is stored in the first of two photorefractive crystals. A second plane wave reference or read beam is diffracted by this grating and reconstructs the phase conjugate of the original object wave. This phase conjugate is interfered with a third plane wave reference and the resulting interference pattern is stored in the second crystal. A fourth reference wave reconstructs the phase conjugate of the phase conjugate, which is just a reconstructed version of the original object wave. The reconstructed object wave is directed back to the first crystal to rerecord the original grating. The advantage in using two crystals over one is now obvious. The holographic information can be read out without being lost. Each crystal refreshes the information written in the other crystal. A diagram of the two-crystal memory is shown in figure 4. To read out the information, a beamsplitter is inserted in the cavity formed by the two crystals, as depicted in the figure.

The above discussion is a simple description of how the two-crystal optical storage device works. However, to fully appreciate how the device works, and some of the potential difficulties encountered in developing an implementable design, a more detailed discussion of the theory and some preliminary experimental results will be presented.

### Theory of Operation

Because the photorefractive crystal generates a phase conjugate of the original object beam, it is referred to as a phase conjugate mirror or PCM. The reflectivity of the phase conjugate mirror is defined as the ratio of the intensity of the phase conjugate wave to the intensity of the incident object wave. Using a four-wave mixing geometry, it is possible with proper beam geometries to have reflectivities greater than 1.0 (ref. 4). As a result, it is possible to build optical resonators using phase conjugate mirrors in place of conventional mirrors. The optical storage device is a type of phase conjugate resonator. The image information, in the form of a complex wavefront, oscillates between two photorefractive crystals or phase conjugate mirrors. In the literature, the term double phase conjugate resonator is used to describe this type of oscillator.

Modelling the behavior of phase conjugate resonators has been the subject of intense research over the last ten years (refs. 5-7). Phase conjugate resonators differ from conventional resonators in some important ways. Two which are relevant to the optical storage device are; 1) the frequency of the light will shift or detune in a two-crystal oscillator to compensate for phase mismatches at the boundaries, and; 2) the structure of the light beam in the transverse direction, or the direction normal to propagation, is not subject to the same constraints as in the case of an optical resonator formed by two mirrors. First, the effect of frequency detuning on the design of the optical storage device will be addressed. According to theory and experiments, the frequency detuning is a function of the cavity length and can be changed by adjusting the length of the cavity (ref. 7). No frequency shift takes place provided that the cavity length is set so that the phase after one roundtrip is an integral multiple of  $2\pi$ . However, maintaining this exact distance over time is difficult. In the optical storage device, a second feedback loop is provided to maintain the phase match if so desired. The total light output of the cavity is detected with a photodetector, the output of the

photodetector is phase shifted, and the phase shifted signal is used to drive an acousto-optic modulator which amplitude modulates the write beam on the first crystal. By amplitude modulating the write beam, the phase of the phase conjugate wave can be varied and adjusted. In this way, the phase match at the boundaries can be maintained. The second difference stated above does not directly affect the design of the storage device, but is the fundamental reason why a phase conjugate resonator can be used to store holographic images. With conventional oscillators, where a beam of light oscillates between two mirrors (as in a laser cavity), only certain structures are allowed in the transverse direction. That is, because of the boundary conditions, and the conditions required for oscillation, the structure of the beam, or intensity profile, can take on only one of a set of distinct patterns. These patterns correspond to what are known as the transverse modes of the resonator (ref.8). When the mirrors of the conventional resonator are replaced with phase conjugate mirrors, the boundary conditions change. Theory predicts that the transverse modes are no longer restricted in form. Therefore, a very complex form or structure, or multiple transverse modes (of the form predicted for a conventional resonator) may be supported simultaneously within the cavity (ref. 6).

The structure, or lack of structure, in the transverse direction predicted for the phase conjugate resonator means that a wavefront of arbitrary phase and amplitude profile can be maintained in the resonator cavity. Therefore, theoretically, holographic image information can be stored as a complex wavefront in the cavity formed by two phase conjugate mirrors. This is the theory behind the operation of the optical storage device.

### Experimental Results

Experiments have been carried out to study the image information storage capability of a two crystal oscillator (ref. 9). In figure 5 is a picture of the optical setup. The smaller photograph, a picture of the Airforce Resolution chart, is the image output from the cavity of the two PCM resonator. For this particular setup, the gain of the cavity was slightly less than one. As a result, the image could not be retained for an indefinite period of time. Rather, the image was stored for approximately a minute. This represents more than a factor of 100 increase over the length of time the image could be stored in one crystal for the geometries and power levels which were used.

When the power levels and beam ratios were set such as to achieve a gain in the cavity in excess of 1.0, the transverse structure of the beam was corrupted for some of the experimental runs. A particular (repeatable) pattern emerged. This was not anticipated. According to theory, as described above, the transverse modes of a PCM resonator are degenerate, meaning that the energy of all transverse modes is the same. This being true, no one mode or combination of modes should dominate. Further experiments and analyses are required to explain the observed phenomenon.

The feedback described above for amplitude modulating the write beam was tested experimentally. The optical setup is shown in the diagram of figure 6. It was found that the ability to control the phase of the phase conjugate wave was limited by the amplifier which was used to phase shift the light signal. It was possible to affect the phase of the phase conjugate, but it was not possible to stabilize the oscillation with the setup that was used. Further experiments are planned along these lines.

### Conclusions

A design has been proposed for an optical storage device which can store holographic information in two photorefractive crystals. The advantages of using photorefractive crystals is that these crystals provide the capability of recording holographic image information at near real-time rates and promise high storage densities. Potential applications for optical storage devices such as the one described in this paper include storage of holographic image information for distributed sensing and processing. A conceptual design of a distributed sensor and processor for large flexible space structures is described in reference 10. All of the techniques which have been developed for holographic sensing of structures using fixed media can now, with the development of a dynamic holographic recording media, be used to develop sensors which can measure the motion of structures.

The image resolution which can ultimately be achieved using photorefractive crystals is unknown. The resolution is a function of the size of the crystal. Techniques are currently being studied which will allow crystals of larger sizes to be grown. Presently, the crystals are limited in size to approximately 5 mm cubes ( $\text{BaTiO}_3$ ). The growth of photorefractive thin films is also being explored. Thin films are much easier to grow, can be fabricated in much larger sizes, and require less time to grow than bulk crystals.

Using BaTiO<sub>3</sub> crystals, experiments have been carried to evaluate the operation of the two-crystal optical storage device. Thus far, good results have been obtained in the laboratory. An increase in image storage time on the order of 100 has been achieved using two phase conjugate crystals rather than a single phase conjugate crystal. Further studies are needed to refine the model and design for the two-crystal oscillator.

### References

1. Ashkin, A.; Boyde, G.D.; Dziedzic, J.M.; Smith, R. G.; Ballman, A.A.; Levenstein, H.J.; and Nassau, K.: Optically-Induced Refractive Index Inhomogeneities in LiNbO<sub>3</sub> and LiTaO<sub>3</sub>, *Applied Physics Letters*, vol. 9, 1966, pp 72.
2. Chen, F.S.; LaMacchia, J.T.; and Fraser, D.B.: Holographic Storage In Lithium Niobate, *Applied Physics Letters*, vol. 13, 1968, pp. 223 .
3. Feinberg, J.: Optical Phase Conjugation in Photorefractive Material, Optical Phase Conjugation, ed. Robert A. Fisher, Academic Press, New York, 1983.
4. Feinberg, J.; and Hellwarth, R. W.: Phase-conjugating Mirror With Continuous-wave Gain, *Optics Letters*, vol 5, No. 12, December 1980, pp. 519-521.
5. Lam, J. F.; and Brown, W. P.: Optical Resonators With Phase-Conjugate Mirrors, *Optics Letters*, vol. 5, no. 2, February 1980, pp. 61-63.
6. Siegman, A.E.; Belanger, P.A., Hardy, A.: Optical Resonators Using Phase-Conjugate Mirrors, Optical Phase Conjugation, ed. Robert A. Fisher, Academic Press, New York, 1983.
7. Ewbank, M.D.; and Yeh, P.: Frequency Shifts of Self-pumped Phase Conjugators, *SPIE vol. 613, Nonlinear Optics and Applications*, 1986, pp. 59-69.
8. Kogelnik, H. ; and Li, T.: Laser Beams and Resonators, *Applied Optics*, vol. 5, no. 10, October 1966, pp. 1550-1567.
9. Welch, S. S.: Holographic Information Storage Using Two Photorefractive Crystals, *Proceedings of the IEEE Lasers and Electrooptics Society 1989 Annual Meeting, Orlando, FLA, October 17-20, 1989*, pp. 439-440.
10. Welch, S. S.: Data Processing for Distributed Sensors in Control of Flexible Spacecraft, *Proceedings of the NASA Workshop on Computational Aspects in the Control of Flexible Systems, Williamsburg, VA, July 12-14, 1988, NASA TM 101578*.

## Figures

Figure 1. Diagram showing two plane coherent waves interfering in a photorefractive crystal.

Figure 2. Graph showing relationship between light intensity, charge density, electric field amplitude, and index of refraction.

Figure 3. Diagram showing four-wave mixing geometry.

Figure 4. Block diagram of two-crystal memory.

Figure 5. Photograph of laboratory setup of two-crystal memory using BaTiO<sub>3</sub>. The stored image (smaller photograph) is of Air Force Resolution Chart.

Figure 6. Diagram of optical setup for two-crystal memory

Figure 1.

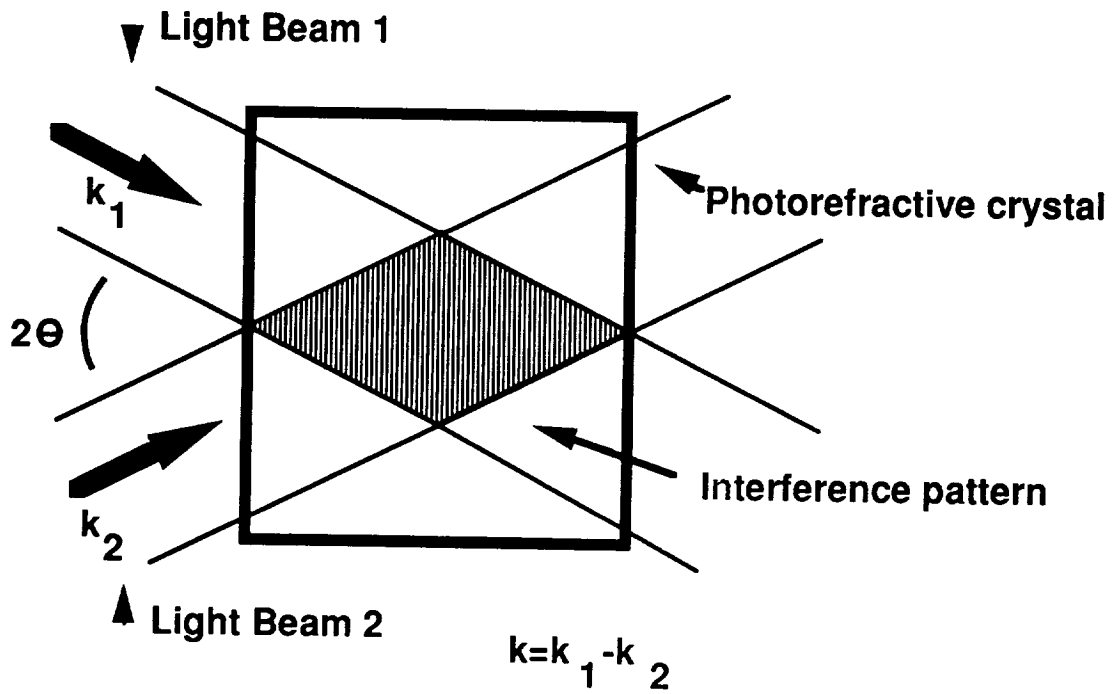


Figure 1.

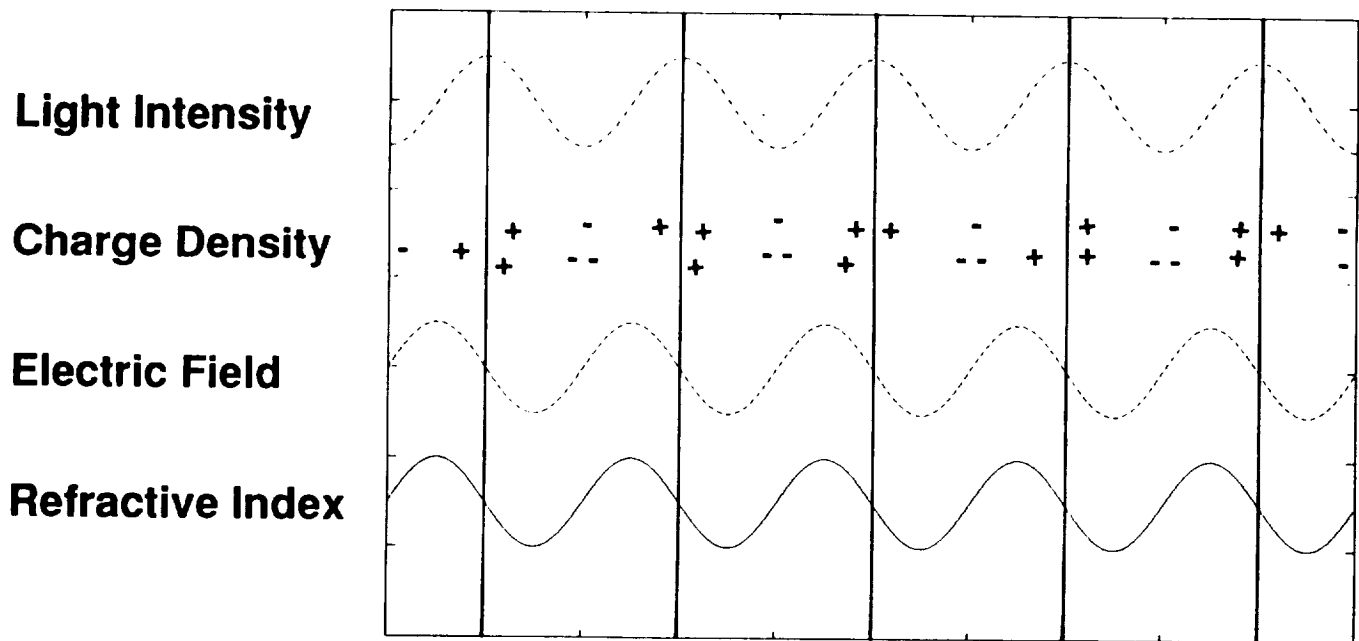


Figure 2.

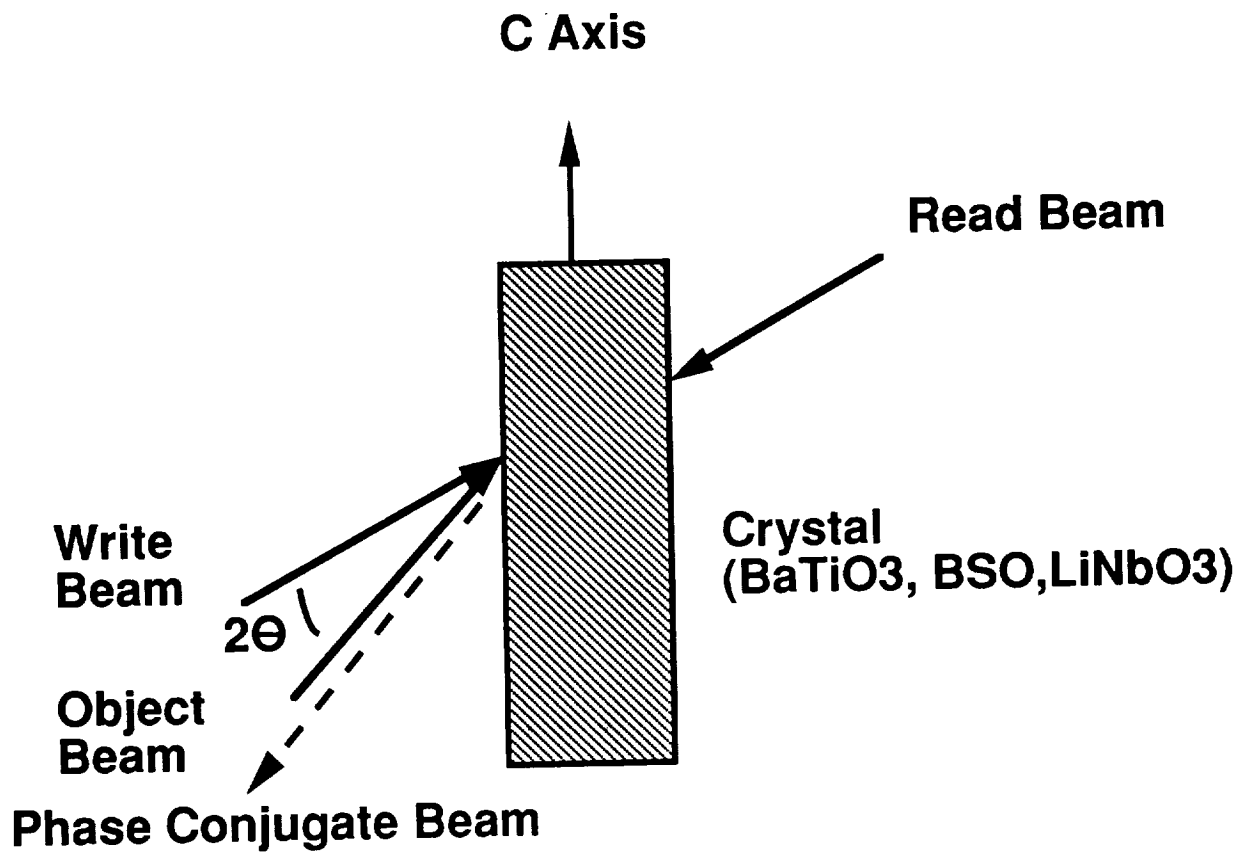


Figure 3.

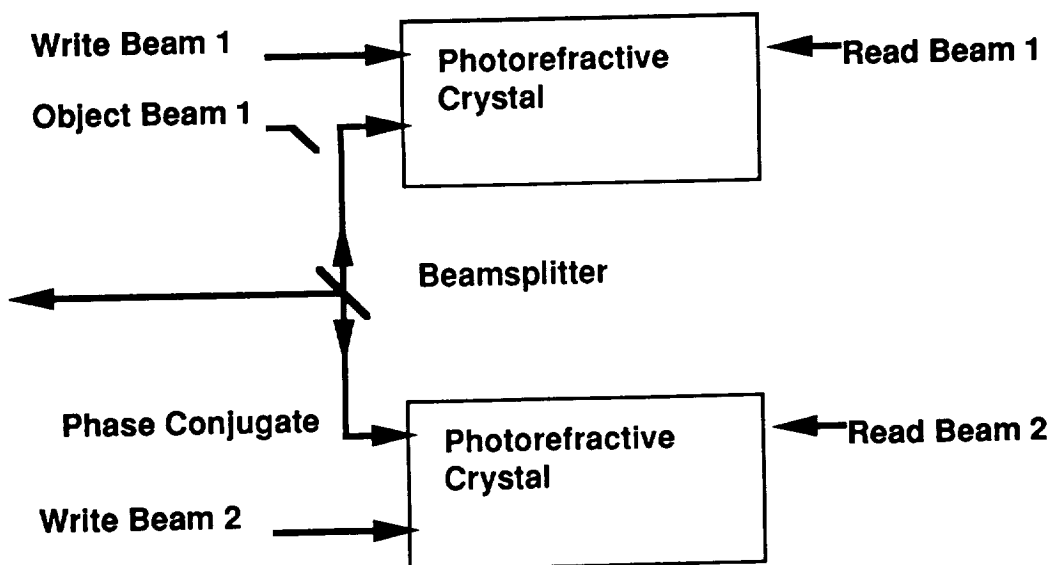


Figure 4.



Figure 5.

ORIGINAL PAGE IS  
OF POOR QUALITY

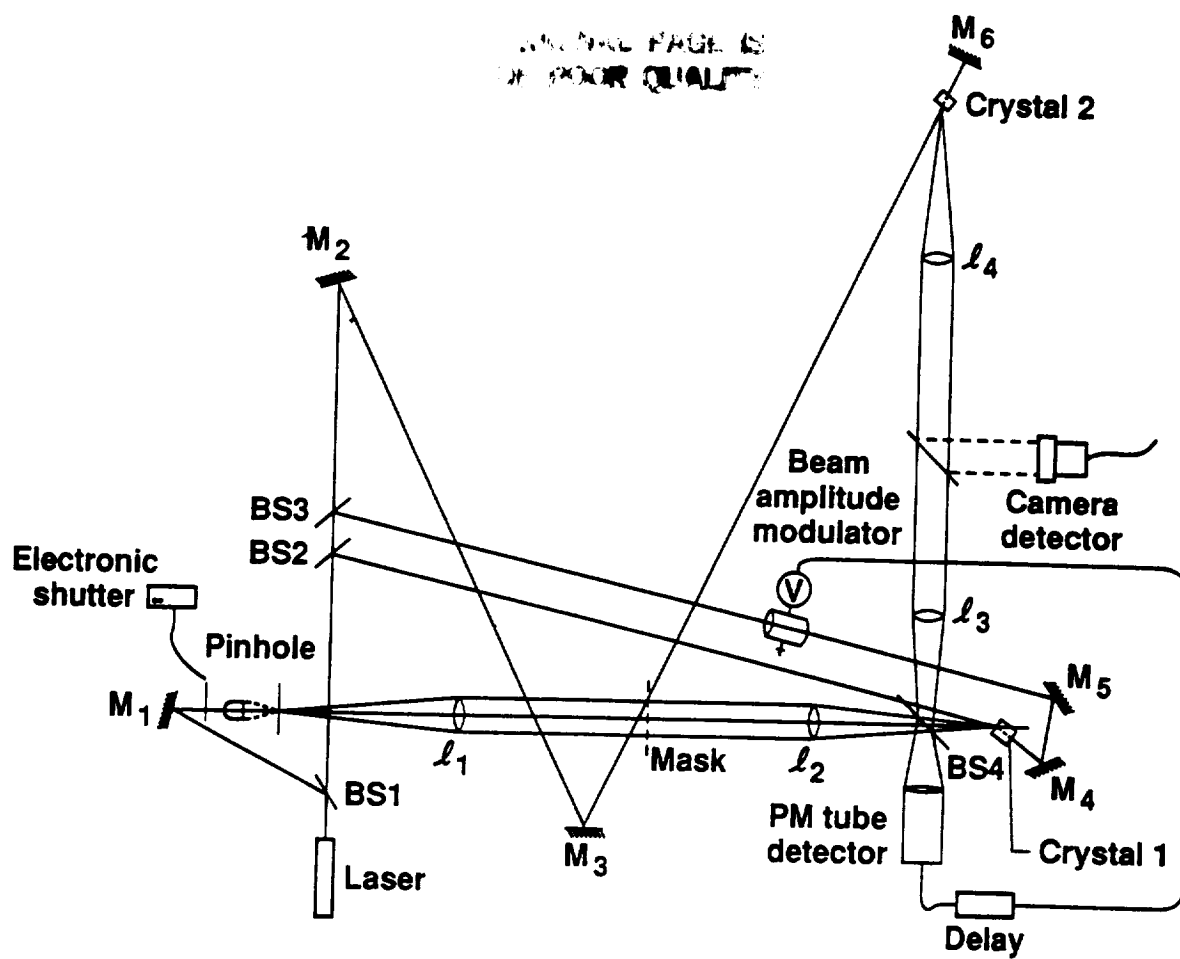


Figure 6.



# REWRITEABLE OPTICAL DISK RECORDER DEVELOPMENT

Thomas A. Shull and Pamela L. Rinsland

National Aeronautics and Space Administration  
Langley Research Center, Hampton, Virginia 23665

## ABSTRACT

A NASA program to develop a high performance (high rate, high capacity) rewriteable optical disk recorder for spaceflight applications is presented. An expandable, adaptable system concept is proposed based on disk Drive modules and a modular Controller. Drive performance goals are 10 gigabyte capacity, 300 megabits/second transfer rate,  $10^{-12}$  corrected bit error rate, and 150 millisecond access time. The design for an expandable Controller is presented. System goals are up to 160 gigabyte capacity at up to 1.8 gigabits per second rate with concurrent I/O, asynchronous data transfer, and 2 to 5 year operating life in orbit. Technology developments, design concepts, current status, and future plans are presented.

## INTRODUCTION

High performance mass storage systems are crucial to future NASA programs such as the Earth Observing System (EOS) polar orbiting platforms. The complexity and capabilities of space information systems continue to grow. The volume of data generated by orbiting scientific instruments is projected to increase by orders of magnitude over the next decade. Requirements for onboard storage with capacities on the order of one terabit ( $10^{12}$  bits) and data rates in excess of one gigabit per second are anticipated. Ground-based mass storage needs such as data downlink buffers, temporary archives, and data processing, particularly by supercomputers and image processing workstations, are also growing.

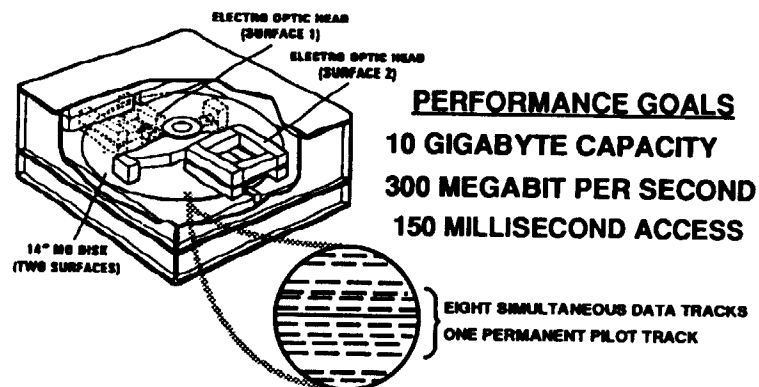
To meet these needs, the Spaceflight Optical Disk Recorder (SODR) program was initiated. This program is sponsored by the NASA Office of Aeronautics and Exploration Technology and managed by NASA's Langley Research Center. The objective of the SODR program is to develop and demonstrate the technology and subsystem elements which form the basis for versatile, expandable mass storage systems for space flight applications. The approach is to produce a stackable, high performance (300 megabit per second, 10 gigabyte) rewriteable optical disk Drive and a modular system Controller. The system goals are 160 gigabyte ( $1.28 \times 10^{12}$ ) capacity and 1.2 gigabit (150 megabyte) per second data rate. The Drive and system concept and corresponding functional architecture are shown in Figure 1. This paper will discuss the concept, plans, status, and applications (including ground-based) of the SODR development program.

## BACKGROUND

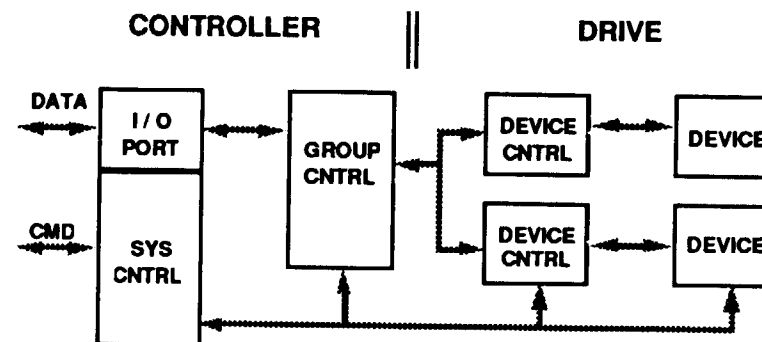
There are three basic types of optical disks. Read only disks, such as CD ROM, are reproduced from a laser written master and cannot be altered. This has become a popular medium for data distribution. Write-once-read-many (WORM) disks allow the user to permanently write information on the disk. These are used primarily for archival applications or when an audit trail is desirable. Both 5.25 (130 mm) and 12 inch (300 mm) write-once commercial products have been on the market for several years, with OEM vendors providing complete (jukebox) systems. Erasable or rewriteable disks allow the user to write, read, erase, and rewrite information. Rewriteable optical Drives, using 5.25 inch disks, are now commercially available. A wide range of information on optical data storage can be found in the SPIE Proceedings.<sup>1,2</sup>

Ruggedized small drives are being developed for the Government. A 5.25 inch write-once Drive is being qualified for a Shuttle experiment by NASA's Lewis Research Center. A 5.25 inch rewriteable disk Drive has been developed and demonstrated in fighter aircraft by the Air Force and is being considered by NASA Langley for aircraft instrumentation programs.<sup>3</sup> This same design has been modified for space flight and Drives are being tested for incorporation into Shuttle experiments by NASA's Goddard Space Flight Center and NASA Lewis.

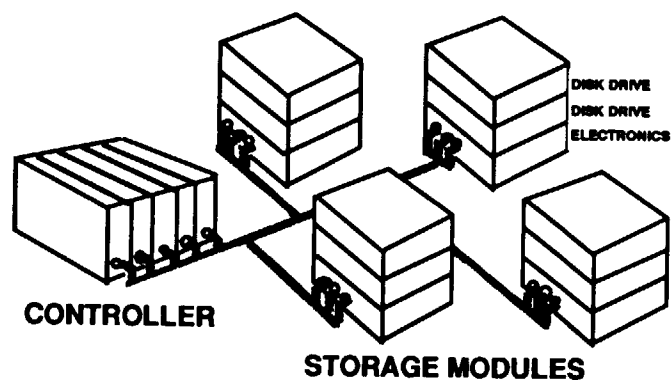
### SODR DRIVE CONCEPT



### ARCHITECTURE (MIN)



### SYSTEM CONCEPT



### ARCHITECTURE (MAX)

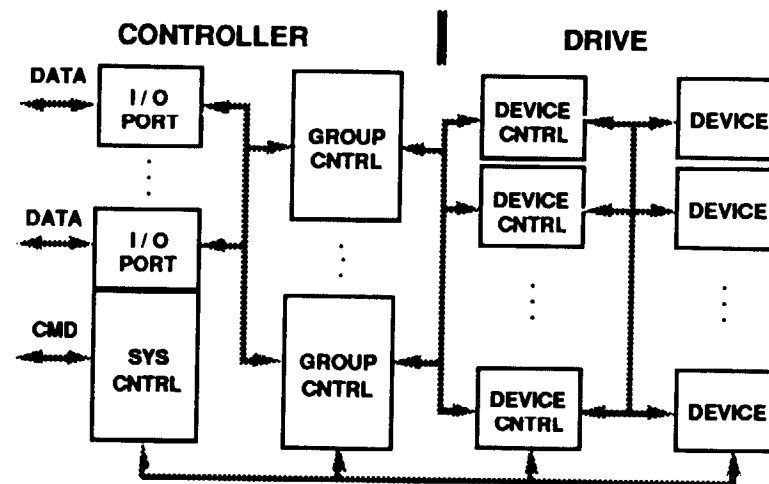


Figure 1. The SODR 14-inch disk Drive and configurable, expandable mass storage system concept. Modular approach based on stackable Drive and modular Controller.

To address high performance needs, the Government has been actively developing 14 inch (350 mm) systems. Two high capacity write-once "jukebox" systems were developed and delivered to NASA's Marshall Space Flight Center and the Air Force. The Air Force is funding the development of a 14 inch rewriteable disk Drive using two simultaneous laser tracks for transport aircraft applications. A consortium of Government agencies, including NASA, sponsored applied research in eight-track rewriteable optical disk storage technology based on a concept known as the Optical Disk Buffer which culminated in a demonstration in July of 1988. Extension of these developments to meet high performance space flight mass storage needs is the challenge being addressed by the SODR program.

## DRIVE TECHNOLOGY

The SODR Drive development is based on magneto-optic (MO) rewriteable optical disk technology. The active media is a magnetic material made from rare earth and transition metal compounds. This is the most popular and mature rewriteable media.<sup>4</sup> The phases of the magneto-optic recording process are shown in Figure 2. The media is initialized to a common magnetic orientation that the inherent magnetic forces retain even in the presence of a strong external field. To write a mark, a focused laser spot is used to heat the media to the Curie point where the external field causes the magnetic orientation to be reversed. The laser is then used as a polarized light source to read the disk. The polarization angle of the reflected light is rotated corresponding to the magnetic orientation due to the Kerr effect. This rotation is detected, indicating the presence of a written mark. To erase the disk, the external field is returned to the original orientation and the laser is used to reheat the media and return the disk to its initial state.

Technology development for the SODR program has been focused on three fundamental areas: ruggedized 14 inch dual-sided MO media; independently addressable nine-element solid state laser diode arrays (used to provide eight parallel data tracks); and demonstration of a multi-track electro-optic (EO) head, with its electronic and mechanical subsystems.<sup>5,6,7,8</sup> Current media studies are focused on optimization of MO performance and the suitability of glass substrates for harsh environments. Tests include vibration, thermal cycling, radiation, and outgassing. A blank glass disk has survived 62 grms random vibration. Preliminary results continue to demonstrate that glass is the preferred substrate over aluminum. The laser development effort is focused on operating life and yield improvements. This includes fine tuning of the structure and process refinement and conversion to MOCVD crystal growth. Techniques to stabilize the lasing frequency, which can shift due to optical feedback and aging, are being studied.

As a proof of concept, a Technology Demonstration Unit (TDU) was produced. The TDU is composed of a multi-track EO head with 9-element diode laser array and a single-sided MO disk. Write, read, and erase of eight parallel data tracks, using a separate laser element for focus and tracking, was demonstrated in early 1990.<sup>11</sup> Figure 3 is a polarizing microscope photograph of actual recorded information (marks) from a MO 14 inch aluminum substrate disk. The effective data rate is 133 Mbps per second. The focused laser spot size is 0.7 microns and the minimum recorded feature length is 0.8 microns; this results in a capacity of 5 gigabytes on a side.

## DESIGN CONCEPT

NASA's investigation of spaceflight applications led to the modular building block system concept shown in Figure 1. The system is made up of multiple Drive units, with their supporting electronics, and a modular Controller. This concept supports a variety of mass storage applications and forms the basis for an expandable system that can be configured for specific applications (capacity and I/O rate). This modular architecture also permits efficient hardware and software development, starting with single modules and isolates internal Drive interfaces from those likely to change, specifically the user interface.

The top level functional architecture is also shown in Figure 1. It is a hierarchical structure, starting with the device as the lowest level and expanding upwards to the system control function. The device performs the basic storage functions: reading, writing, and erasing of information. The device control supports device specific and data processing functions like data formatting, EDAC, and data encoding/decoding. The group control connects devices (through device control) to ports. It provides port related functions such as

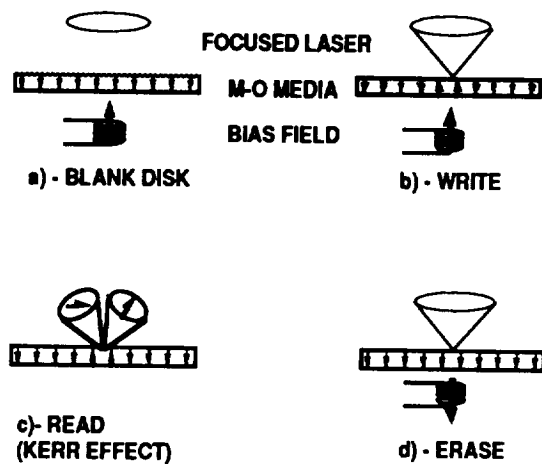


Figure 2. Magneto-optic process: a) initialized media unaltered by external field; b) laser heats media to Curie temperature where external field reverses orientation; c) change in polarization angle of reflected laser light indicates written mark; d) media erased by rewiring to initial orientation.

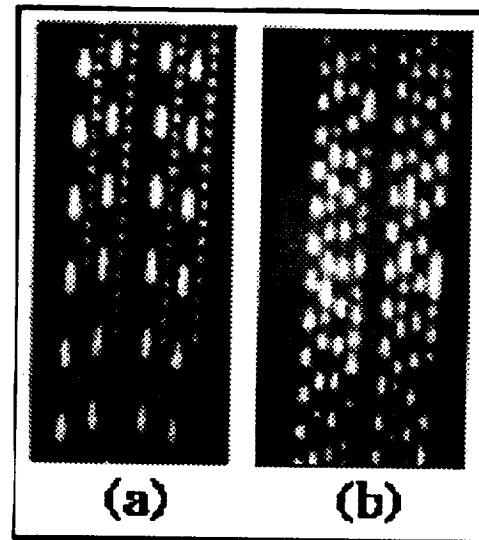


Figure 3. Polarizing microscope photograph of marks written on 14 inch MO disk by Technology Demonstration Unit; a) Alternating 1.875 and 7.5 megahertz square wave; b) Alternating 3.75 megahertz square wave and pseudo random data at 133 megabit per second. Minimum feature is .8 micron. Track spacing is 1.4 micron.

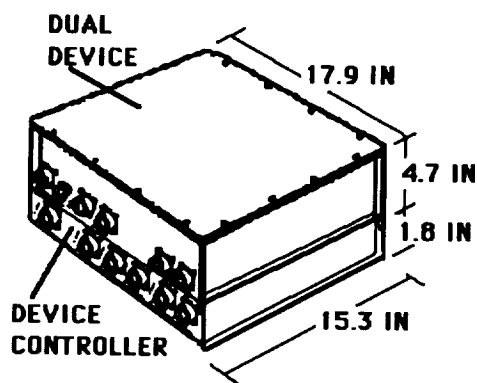


Figure 4. Projected flight Drive packaging with projected size, weight, and power. A Drive contains two devices and separate Device Controller.

#### PHYSICAL CHARACTERISTICS

	DUAL DEVICE	DEVICE CONTRL
SIZE:	3/4 FT	1/4 FT
WEIGHT:	45 LBS	15 LBS
POWER:		
WRITE 150	50W	50W
WRITE 300	70W	100W
READ 150	75W	50W
READ 300	130W	100W
ERASE 300	70W	0W
ANGULAR MOMENTUM:	1 FT-LB	

rate buffering and multiple device synchronization. The system control provides the overall control, command processing, and interface to a single user or network.

The basic functions have been partitioned into a Drive, composed of the logical device and device control functions, and the Controller, composed of the logical group control and system control functions. Ideally, a set of generic control functions and a Controller to Drive interface can be defined which will be suitable for all applications. The detailed functional partitioning of control and system level functions continues. A system model is being developed to support system design and investigate functional issues such as file management, erasure algorithms, resource allocation, and data throughput.<sup>9</sup> This model will be used to validate an expanded multiple Drive system before it is built.

### Drive

The SODR disk Drive unit, shown in Figure 1, is similar to a magnetic or Winchester disk Drive. It contains, as a minimum, the media and its supporting mechanisms. These include the optical heads, rotating disks, photodetectors, and the support subsystems (mechanical, optical, and electronic) needed to read or write information on the media. The SODR Drive contains a single disk with independent multi-track heads accessing each surface. Within the SODR system architecture, the Drive is logically equivalent to two devices. The projected physical package for the Drive and supporting electronics package, which represents the Device Controller, and its associated parameters are given in Table 1 and Figure 4.

The design is based on well known split-head, differential detection MO concepts. In a split-head system the laser and associated collection optics are mounted on a fixed, thermally controlled base plate and only the final objective lens assembly moves. Differential detection means the reflected beam is split into its polarized components and directed to separate detectors whose outputs are compared differentially. Supporting subsystems include the spindle and linear translator servos, which position the head, the data modulator/encoder which converts digital data into analog signals to drive the laser, and the demodulator/decoder used to convert detector system analog signals into digital data.

### Controller

The Controller provides overall system control and the necessary interface between the user and the storage devices. It is responsible for command processing, multiple Drive synchronization, data buffering, file management, logical to physical mapping, fault processing, system self test, status reporting, and user data connection via high speed data ports. As shown in Figure 1, the System Controller will provide command and control operations, and the Group Controller with the Data I/O Ports will provide high speed data processing. The Drive provides data encoding and EDAC. The command and data interfaces are separated to achieve the desired reconfigurability and data throughput. The SODR Controller provides variable data transfer rates, simultaneous input and output through separate data ports, and dynamic reconfiguration.

## CONTROLLER SUBSYSTEMS

The modular design approach has been followed within the Controller by further partitioning into subsystems and modules as shown in Figure 5. Each of these subsystems will be developed separately for the breadboard controller using commercially available technology; eventually they will be packaged as one system element for the flight SODR Controller.

A Group Controller internal circuit module is referred to as a slice and consists of the hardware necessary to implement the complete data path circuit for an eight bit parallel increment of data. The slice concept is consistent with the current design of the storage Device which will record the data on the disk surface in an eight bit parallel format. Each slice can be processed and transferred at a 150 Mbps rate. Data rate capabilities can be expanded by adding one slice for each 150 Mbps increment of the required data rate.

The Data Ports provide modularly expandable user interfaces. To achieve the rate requirements, the proposed ANSI High-Performance Parallel Interface (HPPI) standard was selected to be used for the User

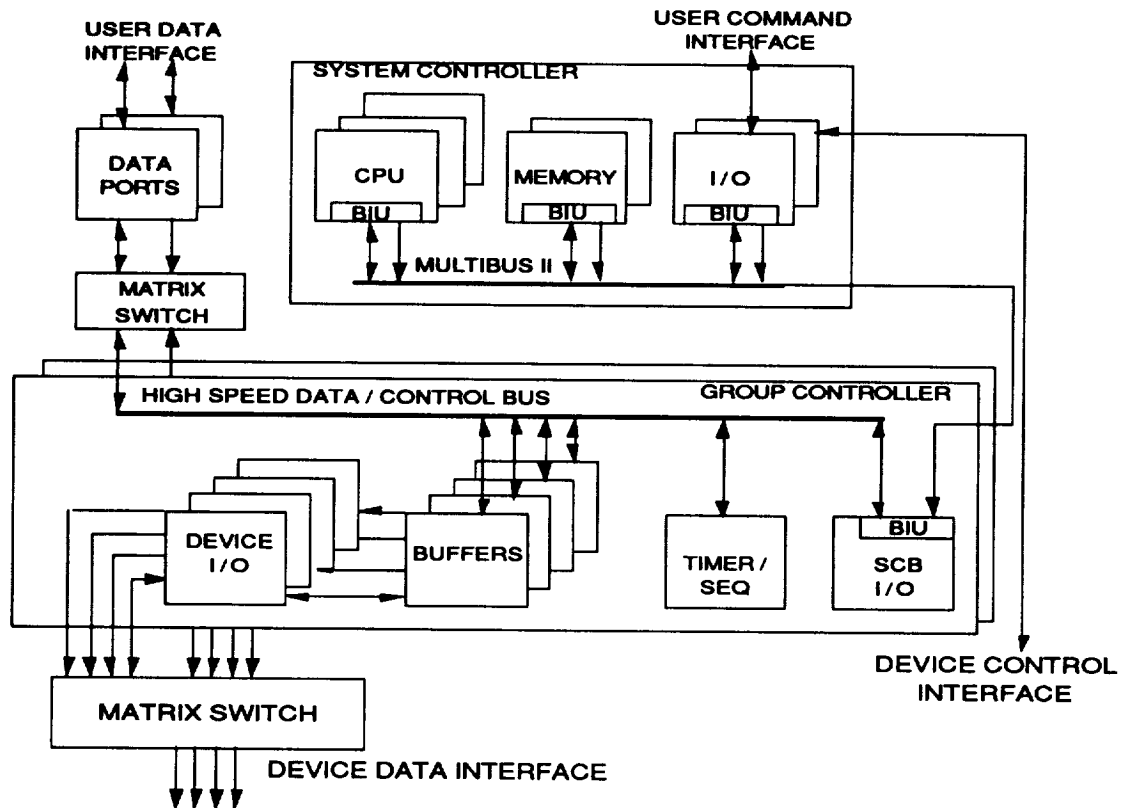


Figure 5. SODR Controller block diagram. Modular design approach which supports multiple Drives, multiple ports and byte wide expansion in data transfer.

interface Data Port. It will be implemented in eight bit parallel modules that correspond to a Controller slice for Drive compatibility and expansion.

The Matrix Switch provides the capability to quickly switch the data transfer path from one subsystem to another. This supports dynamic reconfiguration of the system according to user rate and capacity requests and supports graceful degradation by removing failed components from the data path. It also provides the capability to place unused drives on standby for reduced power consumption. The capabilities provided by the Matrix Switch are primarily for the requirements of flight versions of the controller. Ground-based systems could use direct connections for variations in system configurations.

#### System Controller

The System Controller decodes user commands and maintains control of the rest of the SODR system. The breadboard version will be composed of triply redundant CPU cards, memory cards, user command port cards, and the Multibus II message passing bus as shown in Figure 5. The CPU is the Harris RTX 2000. This processor was chosen for the high speed achieved by its RISC architecture, low power, ease of interfacing and programming, and availability as a standard cell in the Harris HCMOS process.

For the breadboard, the Multibus II standard has been selected as the primary system control bus. It provides operating flexibility and supports simple implementation of maximum system configuration. The Multibus II provides clearly specified electrical and mechanical systems, protocols for message passing, shared memory access, and support for unsolicited messages (i.e. virtual interrupts), and uses a parallel bus for fast data throughput. Also, it allows concurrent operations, permits convenient implementation of system fault tolerance, and is being considered by other projects for space flight use.

The primary functions of the System Controller software are to maximize data throughput, recover from failures, and maintain data file directories. This software does not directly handle the high speed data being transferred across the data port. It acts as a traffic cop for the data by decoding user commands, coordinating resources, and issuing sub-commands to the various hardware elements at its disposal.

A goal for the software design is to exploit the multiprocessor architecture to provide maximum fault tolerance and throughput. This will be accomplished by developing a data directed programming kernel that executes on each CPU. The scheduling of tasks is coordinated with the other active CPUs and results in a dynamic allocation of tasks to CPUs. Dynamic allocation of tasks enables load balancing of CPU resources and automatic recovery from CPU failures. A task is initiated on an idle CPU when the kernel detects that the appropriate arcs are pending for that task and broadcasts this data to each CPU by an unsolicited message. Message handlers on the other CPUs decode the message and start a timer so a recovery routine can be executed if the task is not finished by a specific amount of time. Tasks broadcast new arc data over the message passing system so all CPUs can update their arc and task definition tables.

### Group Controller

The primary function of the Group Controller is data processing, synchronization, and buffering necessary to transfer high speed data between the user and the storage modules. Other functions performed by the Group Controller include System Controller data transfer, self test, and data interface control.

A block diagram of the functional concept for the Group Controller is shown in Figure 5. The Controller is divided into components according to functional requirements. The Buffer provides memory management, configuration control, and dual port RAM storage for data rate synchronization. The high speed Sequencer provides timing, address control, and data routing. The Data Port provides the user high speed data interface. The Device I/O provides the data interface between the buffer and the storage modules. The System Controller I/O provides command and system data transfer between the Group Controller and the System Controller. The High Speed Data Bus provides the interface for high speed data between the Buffer and the Data Port, and the interface for system data and control between all the components of the Group Controller.

One of the areas critical to the success of the Controller is the transfer of high speed data across backplanes and interface cables. Current commercial backplane interface standards have typical bandwidths around 10-16 MHz. Transfer rates are limited by IC propagation data delays, backplane clock skews, and multiple card loading. Backplane transmission line characteristics must be considered to minimize undesirable reflections. Analysis of the Futurebus indicates a best case transfer rate of 23.8 Mbps for synchronous and asynchronous systems. Due to the limitations of these commercial standards, a custom interface design is required for the Group Controller High Speed Data Bus. This custom interface is expected to achieve slightly better performance than the Futurebus standard by implementing the interface with minor variations from the standard. These changes include limiting the card loads to six and bus lengths to six inches and, if necessary, using ECL instead of ASTTL logic for the data latches. Fiber optic technology will also be considered for the design of the flight version of the Controller.

### Interface Ports

The Controller will contain interface ports, as shown in Figures 1 and 4, to provide control and data to the User. The ports are separated into User Command and User Data segments to achieve desired reconfigurability and throughput. An SODR system may contain one or more data ports with slightly different configurations based on the number of slices needed to meet specific applications. The User Data Interface must be easily adaptable to variable maximum data rates; each slice in the User Data Interface will transfer data to and from the user at rates of up to 150 Mbps. The User Command Interface will provide transfer of all commands and status needed by the user to control the data recording operations.

The current User Data Interface design is based on the HPPI standard. Although this standard has not been finalized, available specifications clearly indicate that the basic HPPI concepts are ideally suited for this interface. The 200 Mbps transfer rate for each eight bit segment in the HPPI cable provides a comfortable margin above the system requirements of 150 Mbps for each slice. Design time will be minimized by taking advantage of the specifications and signal protocol within the standard. Also, future commercial HPPI hardware could be adapted to later versions of SODR.

The User Data Interface for the breadboard system will be implemented with currently available technology. The primary functions will be provided by high speed PLD chips and FIFOs, as well as standard high speed logic chips. These designs will be readily adaptable to take advantage of future developments in VLSI and ASIC technologies when the flight version is built.

Current plans for the User Command Interface include the use of standard interfaces such as the RS232, RS422, or Mil-Std-1553. The selected interface plugs directly into the System Controller MultiBus II backplane and minimizes the complexities of changing interfaces to meet different user needs.

#### Drive Interfaces

The actual design of the Device Data and Device Control interfaces has not been started, but the concept and approach have been fully developed. The eight bit parallel slice concept will be maintained at the Device Data interface. Data transfer across this interface will be slightly over 150 Mbps/slice to allow the transfer of both the original data and the associated overhead while maintaining a 150 Mbps delivery rate of the original data. If possible, the Device Control interface will be selected from available standard interfaces that provide the necessary device and data transfer control. The final selection of this interface will depend primarily on the final design details of the SODR Controller and the Storage Drive unit that is being developed separately.

#### DEVELOPMENT PLANS

The SODR Drive development has been divided into three phases. The first phase includes preliminary design of the flight Drive; detailed design, fabrication, and test of a Brassboard Disk Drive; integration and test with a Government furnished breadboard Controller; and verification of both the technology and design by a flight experiment as a Space Shuttle HitchHiker payload. Phase one will address developmental issues associated with the fundamental storage mechanisms specifically focused on spaceflight applications. Some of the issues to be addressed are bearings, bit error rate improvement by EDAC incorporation, flight optics design, flight packaging, and reliability.

The second phase includes design, fabrication, and test of a ground-based, fully functional Engineering Disk Drive (EDD) providing complete 16 track data processing electronics. The EDD will be integrated with a Government furnished Controller to produce a minimum system. This will be installed into a data systems test bed at a NASA facility for demonstration of simultaneous input and output of processed (corrected) data at 150 megabits per second or 300 megabits per second unidirectional data transfer. This unit is most likely to have potential commercial ground-based applications.

The final phase is design, fabrication, and test of a Prototype Disk Drive (PDD). This will include complete VLSI development and environmental testing. The PDD is to be integrated with a Government furnished flight qualified Controller and the EDD to provide a complete two disk system for demonstration and test. At the completion of phase three, a flight qualified Drive should be available for integration with a flight experiment.

#### SYSTEM APPLICATIONS

The random access capability of a disk memory makes many spaceflight applications possible, in addition to those classically served by tape recorders. Applications range from communications link buffer or temporary storage (much like a magnetic tape recorder) to single instrument or multi-user mass storage



subsystem (random access file storage). Random access storage enables data selection for quick-look or priority downlink. It allows onboard data processing or data compression. Operational telemetry data could be stored and analyzed onboard, or a subset transmitted and additional data transmitted as required. A multi-port system with rate buffering enables dynamic experiment operations. The inherent ability to play data back in a first-in-first-out mode increases the efficiency of ground-based data processing over the current tape approach that uses reversed playback to extend tape and magnetic head life.

The EOS polar orbiting platforms have been selected as candidate users in order to focus the program. The SODR has applications to many future programs, such as shuttle payloads, Space Station Freedom, Mars Rover, and other polar and geostationary orbiting platforms. The unmanned EOS has been identified as a system that could greatly benefit from the new technology and is a suitable target in terms of physical and operational environment. It is also on a schedule consistent with proposed SODR development.

Unfortunately, high performance random access storage represents a new system technology for spaceflight. Today's data systems are designed around magnetic tape recorders, a "proven" technology meeting past and present system needs (maybe not future needs). System designers and users are biased in their thinking by the constraints of sequential access storage. Until the optical disk is demonstrated to be a proven spaceflight technology, their thinking will probably remain constrained. The full random access potential of an optical disk system may not be achieved in the initial applications. As the potential advantages of rewriteable optical disk recorders are recognized, total system concepts and design requirements are expected to change.

#### Flight Environment

The space environment imposes unique operational and physical requirements when compared to ground-based or even airborne optical disk applications. Weight, volume, and power are always critical in space, but reliability, self-test capability, commanded or autonomous reconfiguration, and modular packaging are also important. Fortunately, EOS only requires operation on orbit which is a relatively benign mechanical environment. However, launch survival remains a major challenge. Other factors such as operation in a vacuum, zero gravity, subatomic and electromagnetic radiation, and angular momentum or gyroscopic effects associated with the large spinning disks must also be considered.

The SODR architecture and design addresses many of these problem areas. Redundancy, built-in test, and automatic reconfiguration to map out failed elements are planned to provide the graceful degradation needed for extended missions. A packaging concept based on counter rotating disks is proposed to minimize instability caused by angular momentum. To reduce outgassing effects caused by operation in a vacuum, SODR will use materials that do not dissipate or redeposit on optical surfaces. Special lubricants and sealed or magnetic bearings are under consideration for moving parts. Heat transfer problems apply to both the Drive and the Controller. The heat must be removed through the housing by conduction or radiation and the thermal expansion of materials can drive the design of optical systems. This highlights the importance of low power components such as CMOS and VLSI electronics and efficient solid-state lasers.

#### Ground-based

Although the current NASA program is focused on spaceflight, there are comparable ground-based applications that could benefit from an SODR like system.<sup>10,11</sup> The scheduling algorithms developed to handle the separate read, write, and erase cycles to support prioritized and/or shared data access by flight experiments are equally applicable to ground use. Ground configurations can be more elaborate because the typical flight constraints of power, weight, and volume need not be met. One use is as a rate buffer to provide initial storage of telemetry data prior to dissemination to slower speed distribution networks or facilities. As a supercomputer peripheral, a high performance optical disk system would enable more efficient use of these highly capable machines. Since the HPPI interface selected for SODR data transfer is under design and review by representatives from the leading computer manufacturers in the country, it is expected that SODR can be readily matched to one of these powerful CPUs. SODR can also perform in an

archival fashion and provide effective management of large continuously updated data bases. These applications are also being considered in the current program and are viewed as potential spinoffs.

## CONCLUSIONS

High performance rewriteable optical disk memory with random access capability is an enabling spaceflight technology. Utilization of multiple data tracks, multiple disks, and an expandable Controller offers both high data rate and high capacity storage. The technologies have been demonstrated and the plans are in place for a versatile spaceflight mass storage system to become a reality. The spin-off potential for ground based applications is great.

## REFERENCES

1. Knight and Kurtz, Editors; Proceedings of the Optical Data Storage Topical Meeting, SPIE 1078, Los Angeles, California, January, 1989.
2. De Haan and Tsunoda, Editors; Proceedings of the IEEE/LEOS Optical Meeting on Optical Data Storage, SPIE Proceedings to be published, Vancouver, Canada, March, 1990.
3. Naegele, T., "Optical Disks: Taking Flight, Taking Charge," Military & Aerospace Electronics, April 1990.
4. Freese, R.P., "Optical disks become erasable," IEEE Spectrum, February, 1988.
5. Carlin, D.B., "A ten-element array of individually addressable channeled-substrate-planar AlGaAs diode lasers," IEEE Journal of Quantum Electronics QE-23, No.5, May 1987.
6. Edwards, J., "14" Advanced Erasable Optical Disk Development," presented at the IEEE/LEOS Topical Meeting on Optical Data Storage, proceedings to be published, Vancouver, Canada, March 5, 1990.
7. Hockney, R.L., et. al., "Magnetic bearings for a high-performance optical disk buffer," SPIE 899 Optical Mass Data Storage, 46-53, Los Angeles, California, 1988.
8. Levene, M. L., "High performance optical disk recorder - preliminary test results and spaceflight model projections," presented at the IEEE/LEOS Topical Meeting on Optical Data Storage, proceedings to be published, Vancouver, Canada, March 5, 1990.
9. Hines, et. al, "Application of the Architecture Design and Assessment System for the Modeling of a Spaceflight Optical Disk Controller," 22nd Southeastern Symposium on System Theory, Cookeville, Tennessee, March 1990.
10. Dalton, J.T., "Telemetry data storage systems technology for the space station Freedom era," International Telemetering Conference, October 30, 1989.
11. Wallgren, K. & Michael, G., "Storage for supercomputers - a limiting issue," SPIE 899 Optical Storage Technology and Applications, 292-295, Los Angeles, California, 1988.

## Monitoring and Analysis of Data from Complex Systems

Thomas Dollman and Kenneth Webster  
National Aeronautics and Space Administration  
Marshall Space Flight Center

As flight systems become more complex and longer-lived, it becomes increasingly difficult to monitor and analyze their data. One method being developed to address this problem is the use of ground-based information systems in ways that enable the engineer to relate the data to the system design more readily. Another method is to encode knowledge about the system's operation into the ground computer system itself, to serve as a backup to the engineer's analysis and conclusions. These methods, now being tested at the Marshall Space Flight Center, promise to help maintain the high productivity levels of the telemetry analyst over long periods.

### INTRODUCTION

Complex spacecraft, such as the Hubble Space Telescope (HST), are monitored closely and continuously during flight. The initial mission phase of the HST, the Space Telescope Orbital Verification (STOV), was supported by engineers, scientists and managers at MSFC's Huntsville Operations Support Center (HOSC) and at Goddard Space Flight Center's Space Telescope Operations Control Center. Data originating from the HST is being constantly monitored and analyzed by both computers and humans to determine how well this spacecraft is operating. Fault Diagnosis, Isolation and Recovery (FDIR) is the driver for much of the telemetry monitoring and analysis activities. Ground-based facilities have been receiving and processing over 5,000 telemetry measurements per data-transmission cycle from the HST continuously since its launch in April, 1990.

This paper provides an overview of the computational infrastructure in the HOSC which enables the telemetry monitoring and analysis functions for both spacecraft and vehicles to determine their status and to support corrective actions, when required. Also discussed are some new methodologies and tools that have been employed for on-line monitoring and analyses of telemetry measurements, as well as systems that empower the analysts with off-line tools for information retrieval. To provide perspective on the mainline HOSC computational infrastructure, both historical information and an extensive description of work-in-progress are discussed. These currently used and evolving information systems enable the engineering analysts to effectively perform those functions required for ground-based support of flight systems. Some of the software elements described herein may be applicable to other domains wherein complex systems are being monitored and/or analyzed. All elements are described at various levels of functional detail to enable the reader to decide where the likelihood for such applications might exist.

### HISTORY

In the late 1960's the HOSC consisted of a data acquisition system (Honeywell DF224) and a data presentation system (Burroughs B5500 LP) to support initial Saturn IB launches. The first upgrade, for late Apollo missions, was to the data presentation system. This system was upgraded to a Univac 1108 driving 8 digital display devices, and served multiple users through a video-switching system. The system was somewhat interactive in that display images could be dynamically assigned to any of 8 display channels. Software was mission-dependent and since launches were infrequent, this allowed the software to be altered for subsequent missions. The next upgrade was to prepare for Shuttle missions by using Perkin-Elmer systems driving both 10-channel and 8-channel display systems simultaneously. The major new feature was the ability to define displays dynamically and the assignment and scaling of data to lights, meters, and strip-chart recorders. This enhancement was required for increased launch activities requiring rapid reconfiguration. This system also supported simultaneous real-time presentation and near real-time data recall from data bases. However, definitive data reduction was still relegated to off-line, post-mission data processing activities.

## CURRENT INFRASTRUCTURE

In the early 1980's, design began on the next generation of HOSC data display systems which evolved into the currently implemented Peripheral Processing System (PPS). The purpose of the PPS is to serve as the backbone of the Payload Operations Control Center (POCC) in the HOSC, and it was targeted to specifically support STOV. As an offshoot, the PPS was incorporated into the HOSC's Space Shuttle launch support system as well. The PPS microVAX/Perkin-Elmer based system is networked together on an Ethernet. The Perkin-Elmer systems (which serve as Central Processors - CPs) provide centralized functions, such as data acquisition and decommutation, and the microVAXes (Peripheral Processors - PPs) provide distributed functions, such as data display. In addition to its data display capabilities, the PPS also includes the ability to control payloads from the ground. The primary purpose of the PPS is to provide short-term support for shuttle missions carrying a wide variety of payloads. Thus, for quick turn-around between missions, the PPS was designed to be reconfigurable without changing existing software. Since its inception, the PPS has proved to be very flexible in its support of many types of missions/activities.

The CPs perform centralized functions common to the entire system. The primary function of the CPs are to receive telemetry data in downlink format and provide a pathway for commanding remote payloads. The telemetry data is made available to the PPs in Real-Time (RT) and Near-Real-Time (NRT) modes. NRT data is data from the near past that is saved in data bases on the CPs. The amount of NRT data that can be saved is dependent on both mission and hardware complement. The CPs also monitor the telemetry data for exception conditions that are mission-critical and generate messages that are sent to the PPs to advise the users when exceptions occur. In addition, the CPs maintain and distribute to PPs data base parameters.

The PPs provide distributed functions that are user-specific. On a PP a user can build and modify, in real-time if necessary, free-form displays for viewing his RT or NRT telemetry data. The display capabilities include: tabular fields, X-Y plots, limit sensing, and some graphical representation. When the display is executed, the PP requests data base information from the CPs about each of the parameters to be displayed. Based on that information, it then requests, receives, processes, and displays those parameters from the CPs. In addition, the user may build Special Computations for any special processing that required to be performed on telemetry data. A Special Computation is a FORTRAN program which provides algorithms for processing the data. A package of routines is provided for retrieving data and making any derived parameters available for display. The user may build a Special Computation directly with a text editor or he may do so indirectly with a set of forms that is provided. The user provides a list of input parameters, the FORTRAN algorithms for processing those parameters, and a list of derived output parameters. The forms program then generates the code necessary for retrieving/writing the parameters.

In summary, the Peripheral Processor System provides an effective way to simultaneously provide telemetry information to a number of users in the HOSC. It enables users to build, store, retrieve and modify/display pages to monitor RT and NRT data in columnar and graphical form. However, in order to evaluate new techniques of viewing data (and documentation), a number of different systems were tested both on-line and off-line in the HOSC during STOV.

## NEW ON-LINE SYSTEMS ARE TESTED

Considerable design and engineering knowledge about a spacecraft is documented at various points during its lifecycle. The future availability of this information is important to consider in light of the fact that spacecraft like the HST may require a decade or longer from design to launch, and are utilized on orbit for perhaps two more decades. A life-cycle three decades long or longer, as in the case of the Space Station Freedom, precludes the direct involvement of systems and subsystems designers during much of the operations phases. Therefore, systems which enable or aid the FDIR processes by providing fast, accurate ways to retrieve the specific knowledge pertinent to a specific problem need to be available to help engineers maintain high levels of spacecraft functionality throughout its operational lifetime.

During the STOV, a number of new concepts for telemetry monitoring and analysis were tested in the HOSC. Three of these systems, the HST Operational Readiness Expert Safemode Investigative System, the Vehicle Health and Safety Expert System, and the Thermal Control Subsystem Expert System were implemented in a Knowledge Based Systems (KBS) Testbed, and were operated as an adjunct to the standard PPS functionalities by using a High-Speed Peripheral Processor. Each of the three KBSs provides the user with a unique method for relating telemetry values to the status of the HST.

### High-Speed Peripheral Processor

Data for the KBS Testbed was provided by a High-Speed Peripheral Processor (HSPP), which was interfaced to the Central Processors as a standard Peripheral Processor utilizing configuration-controlled PP software and four Special Computations. The HSPP was implemented as a VAXServer 3300 using VMS and special-purpose interface software elements. The HSPP collected 1400+ telemetry data-items per data-cycle from the CPs and distributed this data over the isolated KBS Testbed network, utilizing TCP/IP, to the three KBSs described below.

### HST Operational Readiness Expert Safemode Investigative System

A reusable knowledgebase shell, the Device Reasoning Shell (DRS), elaborated upon below, was built to access a variety of device models and rule bases to allow a user to solve a variety of problems. To test and demonstrate the application of DRS to the HST, a DRS application was developed which provides assistance in the investigation of HST system anomalies, which are called safemode events [1]. The application, the HST Operational Readiness Expert Safemode Investigation System (HSTORESIS), contains rule bases and a device model that encode knowledge about HST safemode detection, fault isolation, and recovery. A safemode event occurs when any one of a number of sequences of on-board measurements are determined to be off-nominal by the on-board Data Management Subsystem (DMS). The DMS then commands the spacecraft to enter a "safe" or protected mode. These on-board measurements are also available to ground support personnel, so that they can analyze and fix any problem from the ground, prior to uplinking commands to return the HST to normal operating mode. Therefore, it is necessary to thoroughly understand the causes of the sequence(s) of off-nominal measurements which resulted in the triggering of the safemode action. Procedural knowledge was obtained for HSTORESIS from HST safemode system and flight software documentation, and from HST design engineers. The HSTORESIS implementation provides the user with assistance in the investigation of safemode events. HSTORESIS runs as an application using the Device Reasoning Shell (DRS) on a Sun 4/260 workstation.

### Device Reasoning Shell

The Device Reasoning Shell (DRS) is a reusable knowledgebase shell which can access a variety of device models and rule bases to allow a user to solve a variety of problems [2]. DRS was built using Common Lisp and Intellicorp's Knowledge Engineering Environment on a Symbolics 3670 and Texas Instruments' microExplorer before installation on the Sun 4. The DRS is designed to assist in solving problems that require the ability to reason about the model of a device. Knowledge about a device is captured within DRS rule bases and device models. Two programming paradigms, object-oriented programming and rule-based programming, are used to reason about device models. Interaction between the user and DRS is accomplished through a point-and-click style of interface, using buttons and windows. The DRS has three layers: Telemetry Interface, Model Manager, and User Interface. In addition, a data manager records all telemetry items for later replay and analysis, and distributes data items throughout the system.

**The telemetry interface.** The main function of the telemetry interface is to convert raw telemetry into a level of abstraction which is closer to the mental representation that human experts use. This telemetry passes through the telemetry interface and is categorized into one of several different telemetry types to be utilized by the other parts of the DRS. For the HST, 5,500 telemetry items are the main source of information about the behavior of the spacecraft. For the HSTORESIS application, each of the 234 different safemode-specific telemetry data items is mapped into its category (polynomial fit, counter, table lookup, bi-level, or multi-level), and then associated with information about that type.

**The model manager.** The model manager manages Device Models. A Device Model in DRS includes a mapping between the model and a set of monitors, pointers to rule-bases that are capable of reasoning about the device, behaviors that represent the conceptual or physical functioning of the device or component, and features that hold state information that is not included in the satellite telemetry stream. For HSTORESIS, the device model provides a model of the behavior of a device as a function of telemetry data. Devices include physical items like Reaction Wheel Assembly or Solar Array Assembly on the HST.

**The user interface.** The user interacts with DRS through a desktop. The user may point to and click on buttons which activate functions within the DRS application. These functions include startup & shutdown, selection of a message or data-item displayed in a window for elaboration & explanation, moving & resizing windows, and selecting window types for display of telemetry data in different formats. For the

HSTORESIS application, the user can create a number of "monitor pages" or screens-full of data displaying telemetry items per screen-full in formats which include actual value, strip-chart recorder, and/or X-Y plotter displays. Additionally, safemode events detected by HSTORESIS enable access to pop-up windows which describe causes, effects, and recovery procedures relating to the event(s).

#### Vehicle Health and Safety Expert System

The Vehicle Health and Safety (VHS) Expert System was developed at Lockheed Missiles and Space Company (LMSC), using the LMSC-proprietary L\*STAR real-time expert system tool-set. The VHS system was implemented and tested in the KBS Testbed during STOV using a beta-test version of a commercially available real-time expert system package, Talarian Corporation's R\*Time. This software was implemented on a VMS-based VAXStation 3100 and provided a sophisticated monitoring strategy of three different HST flight modes representing start-up (deployment), normal operations, and anomalous behavior (safemode) [3]. The workstation display, utilizing VI Corporation's DV Draw, provided telemetry monitoring graphics as follows.

Four graphs are displayed at the top of the workstation screen which summarize the pointing control and electrical power health of the spacecraft, and allow an observer to spot adverse trends as they develop. Spacecraft mode changes are color-coded, as are checklists of functions to be performed. If some activity is delaying a checklist function, the user may expand the list to observe telemetry values pertinent to that checklist item. In normal mode, display emphasis is on simultaneous status summaries of all pertinent spacecraft subsystems which are required to maintain health and perform the intended mission, the successful acquisition of science data. Safemode is autonomously entered by the spacecraft when any of several dozen reasonableness tests fail. That is, the on-board computer monitoring those sensor readings senses a problem. At that point, the important sensor readings are captured in a recorder, for later analysis by ground station. The ground-based VHS system also captures this data and performs analyses on it, provides status information for each data item, and recommends recovery procedures. The VHS therefore provides the user with a highly flexible monitoring station containing the procedural knowledge related to entering and recovering from anomalous behavior.

A special interface between the VHS Expert System and HSTORESIS was also developed and tested during STOV. The VHS system continuously monitored data pertaining to safemode events, and upon detecting such an event, VHS notified HSTORESIS that an event had occurred. At that point, the HSTORESIS user could access the pop-up windows and then perform an analysis on the pertinent subset of telemetry data. The analysis, via interpretation of strip-charts or X-Y plots, can be performed by replaying, pausing, and "stepping" through any of the previously stored telemetry signals for any portion of the 5 orbits (450 minutes) leading up to that event.

#### Thermal Control Subsystem Expert System

The Thermal Control Subsystem (TCS) Expert System was also developed at LMSC and utilizes identical development and delivery software and hardware as the VHS system [3]. This system was also implemented and tested in the KBS Testbed during STOV, and directly supported HOSC-based thermal engineers throughout STOV. Sensors on the HST enable ground-based FDIR activities on the TCS via some 1400 different data values. These sensors measure the temperatures of the HST internal and external structure as well as mechanical, electronic, and optical equipment (for overheating or overcooling alerts or warnings).

The TCS system contains several hundred solid images of spacecraft subsystems, components, and structures on line which are arranged in a hierarchical fashion. The user can "navigate" through the hierarchy by replacing a current display with a subwindow that is higher or lower in the hierarchy, thereby going from system to subsystem to component in a point-and-click fashion. Each image shows pointers to the location of its sensor(s) in a spatial fashion with real-time readouts of those sensors from the telemetry data. This spatial representation is especially helpful when considering the spacecraft's position with respect to the heating and cooling effects of the space environment. Should a telemetry value achieve an off-nominal value (too hot or too cold), the indicator value changes color depending on the severity. Also changing color are pointers that guide the engineer through the image hierarchy. Displayed on all image levels are warning/alert windows which report on the number, severity, time, and name of the off-nominal sensor(s), and which enable the user to directly access the image where each off-nominal sensor is located.

Additionally, the user can request, in an overlay window, a real-time plot of any sensor readings with respect to time. The TCS Expert System, implemented with R\*Time and DV Draw, provides a robust, comprehensive telemetry monitoring system for the extensive Thermal Control Subsystem of the HST.

#### OFF-LINE AIDS TO ANALYSIS

Information systems which do not directly access telemetry data can also be extremely useful in evaluating that data. Decade-long design/development phases produce a wealth of design and engineering documentation which must be accessed quickly during operations phases. Two approaches to this access are described below. One system, the Document Retrieval Assistant, was utilized during STOV to enable fast access to stored documentation on the design, engineering, and operational aspects of a key part of the HST, the Orbital Telescope Assembly. The second tool, the Design Alternatives Rationale Tool, was developed to capture design knowledge about the Space Station Freedom during the engineering trade-study process, and is beginning to be utilized for fault analyses as well.

##### Document Retrieval Assistant

**Background.** The process of designing, developing, and testing any complex system or subsystem generates a large number of documents to describe the system and processes involved. In the case of the HST, the document-set size and scope is formidable, numbering in the tens of thousands of documents [4, 5]. Documents for programs are typically organized hierarchically, as in a "document tree." Systems are represented as collections of subsystems; subsystems are represented as collections of components. At each level, these elements interact, intercommunicate, and/or interconnect. In order to focus on particular elements or characteristics of a system, to learn or to refresh one's memory, quick access to all aspects of that element in that documentation-set is often required. These documents are often stored electronically, as bit-mapped images and/or text. The challenge is to find all pertinent facts about the subject, including interfaces, in an efficient manner.

**Objective.** To meet this challenge, a system has been developed for MSFC by Hughes Danbury Optical Systems called the Document Retrieval Assistant (DRA). This system serves as a complement to the Hubble Space Telescope Management Information System (TMIS) which contains, on optical discs, bit-mapped images of all drawings, specification documents, and top-level reports and memos relative to the HST. The DRA contains audio interviews, videotapes, and intelligent pointers to TMIS contents as relates to the Optical Telescope Assembly (OTA), a key part of the HST. The DRA was needed to provide the information to support OTA performance analysis and subsystem troubleshooting, repairing and reverifying certain defective sensors returned from orbit, and developing next-generation designs and modifications for certain sensors. The DRA was also developed to capture and file away OTA expert knowledge for retrieval of critical information when required, and for training and familiarization of personnel.

**Using the DRA.** The initial DRA system is fielded on an 80386-based workstation with windows-based user interface and strong emphasis on graphics. The DRA has four main components: hyper-text organization of documents, intelligent search, databases, and system administration and authoring tools. The DRA permits rapid search and retrieval of a very large collection of program documentation (text, photos, drawings, and databases) without resorting to keywords. The two major methods of retrieval are by document browsing (casual or in-depth) and document search (exact text matching, conceptual matching, and fuzzy matching). In short, the DRA is a tool which provides the user with an efficient tool to find all references to a particular component or process, and shows the user drawings of devices or abstracts of pertinent documents, as well as pointers to audio/video discussions and specific documents within the TMIS or other archives for retrieval of more in-depth information.

##### Design Alternatives Rationale Tool

One important aspect of completely understanding complex systems is the need to identify and characterize all possible failure modes. For instance, Failure Modes and Effects Analysis (FMEA) is performed on each of the Space Station's critical subsystems. It is important to ensure that all failure modes have been identified and their full effects are understood, including effects in other subsystems. Much later, in an operational environment, the analyst-team must correctly interpret a fault and understand its consequences before corrective action is taken. One tool, the Design Alternatives Rationale Tool (DART) can be used now by reliability engineers engaged in the FMEA process to build lists of the potential faults that may occur and relate those faults to different failure modes. The methods used in DART for the development, storage, and access of FMEA information should make it useful for the analyst to use during operations.

DART was originally designed for the Space Station Freedom Program as a tool to aid in the engineering analysis process and to easily capture rationale about the alternatives evaluated during its use [6]. It is useful for applications such as trade studies, aiding and documenting decisions made in meetings, and group data gathering. DART employs repertory grids as one form of knowledge representation. This representation appears to the DART user as matrices and allows the critically important capture of rationale behind the axis-entries. DART works well on analysis problems or those problems whose solutions can be enumerated (e.g., classification, interpretation, and diagnosis). Analyses of failures of complex systems fall into this category, so the DART tool is beginning to aid the FMEA process. DART was originally developed on Sun and VAX/VMS computers, but MacIntosh, IBM PC, and IBM mainframe versions are also available.

#### INFRASTRUCTURE EVOLUTION: THE MARSHALL INTEGRATED SUPPORT SYSTEM

An effort is currently underway to replace the current PPS in the HOSC with a state-of-the-art system. However, due to resource constraints and the need to retrain users, it is not feasible to replace the entire PPS in a single step. Thus, it is necessary to follow an evolutionary path where new technology is phased into the old system over a period of time. This allows users to become familiar with new technology while still using the older technology and spreads the cost of development and upgrade over time. Systems such as those described above provide insight and ideas for the evolution of this computational infrastructure.

##### Overview

As a proposed first step along this evolutionary path, the Marshall Integrated Support System (MISS) is being developed to support the Payload Operations Control Center (POCC), a HOSC infrastructure for support of shuttle payload elements. The main requirement of the MISS is to produce, by mid-1991, a state-of-the-art system capable of performing the current PP functions on high-resolution graphics workstations in place of the current DEC microVAXes. Except for data base services, the functionality of the CPs will not be modified until a later date. The system is being designed to support the Space Station Freedom and Advanced X-ray Astrophysics Facility, as well as the current POCC. In addition, it will provide a migration path for utilizing workstation technology throughout the HOSC.

Like the current PPS, the MISS will support many types of missions without the need for software modification. Although the functionality of the MISS is the same as that of the PPS, the MISS is quite different in that it takes advantage of graphics workstation technology. The system is implemented in C under the UNIX operating system and uses MOTIF for its look and feel. The major advantages of the MISS over the current PPS are its primary design goals, modularity and portability.

The functional design of the PPS is very interdependent, whereas, the MISS design is modular, so that each of its components is as independent as possible from all of the other components. In some cases, the components can be used entirely independent of the MISS. In addition, the modularity is further enhanced with packages of routines that are designed to isolate internal and external interfaces and commercial software packages. The modular design makes it easier to integrate enhancements and newer technology into the system. It also helps during development, because it is easier to test the different modules before integrating them with the rest of the system. In addition, modularity makes it possible to integrate commercial off-the-shelf (COTS) products into the system. For instance, if a user in a future mission has a requirement for a COTS product to analyze data, it will be simpler to interface that product to the system.

The current PPS is written in VAX-specific FORTRAN and is practically . The MISS is designed to achieve portability by using industry standards and by implementing the system on multiple platforms. The standards chosen for the system, UNIX, C, X-Windows, and MOTIF were chosen because the graphics workstations initially targeted to become MISS platforms will support them. This helps to enforce portability in that the platform dependency areas can be identified and isolated early in the project.

The functionality of the MISS breaks down into the following subsystems: Data Processing, Data Display, and Payload Commanding. The Data Processing subsystem allows the user to define a set of parameters to be retrieved from the CPs and the processing that is to be performed on those parameters. The Data Display subsystem allows the user to build graphical windows using a wide variety of textual and graphical methods to represent his data. The Payload Commanding subsystem provides a mechanism for the user to control his payload. As indicated above, each of these subsystems is modular.



### Data Processing Subsystem

The Data Processing subsystem breaks down into three tasks: a Data Set Builder, a Data Processing Precompiler, and a Data Retriever. The Data Set Builder is the user friendly windowing interface for defining a set of parameters to be retrieved from the CPs, the processing to be performed on those parameters, and a set of derived parameters to be made available to other processes. The definition is used to create a text file of pseudo code which the Data Processing Precompiler translates into C code. The precompiler then executes the C compiler to compile and link the C code into an executable. This executable receives the raw telemetry parameters from the Data Retriever, converts and calibrates those parameters, executes the user specific processing, and makes the derived output parameters available through UNIX socket connections to the Data Display and the Payload Commanding subsystem. If there are insufficient resources in a single workstation to accommodate all of the subsystems, the socket connection allows the Data Processing subsystem to be moved across a network to another workstation, a further demonstration of the flexibility of the system. The precompiler also creates two additional files: a parameter request list which the Data Retriever uses to request parameters from the CPs and a data set list which describes the derived parameters that are available to the other subsystems.

The Data Processing subsystem combines the Special Computation, type conversion, and calibration functionality of the PPS into a single module. In the PPS data conversion and calibration are a part of the data display capability. In the MISS type conversion and calibration are assumed to be a subset of data processing. The Data Display subsystem deals only with data representation (not with data processing).

### Data Display Subsystem

The Data Display subsystem breaks down into two tasks: the Display Builder and the Display Driver. The Display Builder is the user interface for defining the way, either graphical or textual, that data is to be represented on the screen. The Display Driver makes socket connections to the Data Processing subsystem to acquire data which it uses to update the user's display that he created with the Display Builder. This subsystem is the most flexible in that it is designed so that the Display Builder may be replaced with any appropriate COTS data representation tool. This required that the Display Driver be tool-specific and must be implemented to support the specific tool that is used. The Transportable Applications Environment Plus (TAE+) has been chosen as the default data representation tool. Other COTS products are being evaluated in order to demonstrate the flexibility of this subsystem.

### Payload Commanding Subsystem

The Payload Commanding subsystem breaks down into two tasks: the Command Interpreter and the Command User Interface. The Command Interpreter implements the MISS System Test and Operations Language (MSTOL) which is the language through which users may control their payloads. The MSTOL includes features for modifying and uplinking commands, for accessing telemetry data, and for performing conditionals and loops. The Command Interpreter will accept MSTOL directives interactively or from a script file. The Command User Interface is the user-friendly windowing interface which allows the user to interact directly with the Command Interpreter, query the Command Database, and edit MSTOL script files. The MSTOL represents a significant improvement over the current PPS. The PPS provides mechanisms for modifying and uplinking commands, but does not provide a scripting language for defining logic to control a payload.

### External Interfaces

A main menu interface is provided to allow the users access to the various subsystems within the MISS. The main menu, along with the interfaces to each of the subsystems, use the X Window System as well as the Open Systems Foundation MOTIF user environment to provide a consistent behavior throughout the MISS application. From the main menu, the user has access to a variety of applications including building and driving displays, commanding, building data sets, viewing the telemetry database, and accessing the UNIX operating system. Taking advantage of workstation capabilities, the user may access any number of these applications simultaneously and as often as workstation resources will allow.

Two other PPS components that are being replaced are the Command and Telemetry Databases. Each of these databases are custom, in-house databases which describe the telemetry parameters that are available and the commands that may be uplinked. These databases are being implemented with a commercial Relational Database Management System (RDBMS). RDBMSs allow new information to be added to the databases as it is required thus adding to the flexibility of the new databases.

## SUMMARY

This paper has introduced some of the methods, systems, and prototypes that have been tested for monitoring and analyzing the data from several spacecraft and vehicles at the Marshall Space Flight Center. For the HOSC infrastructure, the MISS provides a migration path to the state-of-the-art workstation environment. Its modular design makes it possible to implement the system in stages on multiple platforms without the need for all components to be in place at once. In summary, the MISS provides a flexible, user-friendly environment for monitoring and controlling orbital payloads. In addition, new capabilities and technology may be incorporated into MISS with greater ease. The use of information systems technology in advanced prototype phases, as adjuncts to mainline activities, is useful to evaluate new computational techniques for monitoring and analysis of complex systems. Much of the software described in this document (specifically, HSTORESIS, DRS, DART, elements of the DRA, and software for the PPS and the HSPP) is available with supporting documentation from the authors, and may be applicable to other systems monitoring and analysis applications.

## REFERENCES

- [1] Cox, P. A.; and Forbes, J. H.: An Application of Design Knowledge Captured from Multiple Sources. Proceedings of the Fifth Conference on Artificial Intelligence for Space Applications, NASA CP-3073, 1990, pp. 291-300.
- [2] Cox, P. A.; and Forbes, J. H.: Hubble Space Telescope Design/Engineering Knowledge-base Device Reasoning Shell (HSTDEK-DRS) User's Manual. LMSC/F373327. Lockheed Missiles and Space Company, 1990.
- [3] Sampedro, P. S.: Hubble Space Telescope Real-time Expert Systems. Unpublished. Lockheed Missiles and Space Company, 1990.
- [4] Clapis, P. J.; and Byers, W. S.: Intelligent Search and Retrieval of a Large Multimedia Knowledgebase for the Hubble Space Telescope. Proceedings of the SPIE Conference on Applications of Artificial Intelligence VIII, vol. 1293, 1990, pp. 590-593.
- [5] Clapis, P. J.: Optical Telescope Assembly Project Report: OTA Document Retrieval Assistant Prototype. PR-1423. Perkin-Elmer, 1989.
- [6] Boose, J.; Chew, S.; Shema, D.: DART User's Guide (Draft). Boeing Computer Services, 1990.

# N91-23036

## HIGH DATA RATE SYSTEMS FOR THE FUTURE

John Chitwood

Goddard Space Flight Center  
Code 727

### BACKGROUND

We are truly living in the information age. Information systems in the next century will transfer data at rates that are much greater than those in use today. The reason for increasing data rates is simple. All of us will be using sophisticated communication devices that require large amounts of data sent to and from them, in as small of an amount of time as possible, in order to maintain the user friendliness of these systems.

Some of these systems of the future include high resolution multi-channel television, high speed facsimile, high speed videotext, and large volume data dissemination services. Business will use such media for electronic mail, on-line libraries, financial reports, and computer forums. The home user will benefit from these services for entertainment, interactive shopping, and library services that include electronic newspapers, publications, periodicals, and encyclopedias.

The users of these systems will be located in high density urban areas, rural areas, and in some cases very isolated, remote areas. Independent of where the user is located, it will be necessary to network users in these different areas. Satellite based communication systems will play an important role in networking these users, whether it be from a user in one urban area to a user in another urban area, or whether it is from a user in a more populated area to a user in a very isolated, remote area.

A typical satellite based communication system is shown in Figure 1. While data rates between a satellite and ground station can be expected to be very large, the data rates between satellites can be expected to be even larger. In addition to this scenario that will provide transmission of data used for commercial communications purposes, NASA has its own high speed data communications requirements. Figure 2 represents the use of tracking and data acquisition satellites to deliver science and engineering data from various scientific spacecraft to ground stations. These tracking and data acquisition spacecraft will be connected together by "crosslinks" to enable the system to maintain near continuous communication with low earth orbiting scientific spacecraft. Again, the data rates that are transmitted between the tracking and data acquisition spacecraft may be quite large.

### SOME TYPICAL DATA RATES

For NASA's data transmission requirements, studies have been conducted on spacecraft that will be launched in the year 2000 and beyond. The spacecraft that have high data rate requirements all have earth imaging instruments on board. The raw data rates that these instruments are expected to output range from about 700 megabits per second (Mbps) to 1300 Mbps. Other studies have been conducted to estimate the data rates that will be required for the tracking and data acquisition satellite to tracking and data acquisition satellite "crosslink". Most of these studies indicate that data rates of 2 gigabits per second (Gbps) will suffice for crosslink applications through the year 2010.

Similar studies have been conducted for commercial communications applications. Some of these project that data rates of 600 or 700 Mbps will suffice for the intersatellite communications link, while others project that a rate of 2 Gbps is more realistic. When one considers that a single high definition digital television signal can require transmission at 650 Mbps or higher, the 2 Gbps estimates could be extremely conservative.

### MICROWAVE, MILLIMETER WAVE, OR OPTICAL

Microwave frequencies, that is those below 30 gigahertz (GHz), can be considered for high data rate transmissions. As the microwave band that could support the highest data rate is only about one GHz wide, this would translate into its only being able to handle data at rates up to one gigabit per second using conventional (quaternary phase shift keying) techniques. Another serious drawback is that these frequencies are allocated for services other than space communications. Radiolocation and radionavigation services represent potential sources of interference for systems using these frequencies.

Millimeter wave frequencies, those above 30 GHz, are suitable for high data rate communication systems. There are frequency bands that have been allocated exclusively for earth-to-space, space-to-earth, and space-to-space applications. There is an additional advantage to using some of the frequencies within the millimeter wave region. Figure 3 illustrates the typical attenuation through the earth's atmosphere for frequencies from 30 to 250 GHz. Some frequencies, such as those around 60, 120, and 180 GHz exhibit an extremely large amount of attenuation through the earth's atmosphere due to oxygen and water vapor absorption. By using frequencies that have been allocated around these absorption bands, space-to-space communication systems may be designed which will have a near zero probability of being interfered with from sources on the earth.

Optical communications technology is also very attractive for high data rate communication systems. Extremely large amounts of bandwidth are

available to serve practically any application. Earth-to-space and space-to-earth optical communication systems present great challenges to optical communications technology. The presence of clouds or fog, and other atmospheric phenomena, make communications through the atmosphere almost impossible except under the best of conditions.

There have been numerous studies conducted that have compared millimeter wave with optical communication systems. The results of these studies have often been biased depending upon the organization that has funded the study.

In general, these studies have indicated that millimeter wave systems have advantages over optical systems in that millimeter wave technology is more mature, and that pointing and tracking requirements for millimeter wave systems are not as critical as for their optical counterparts. Optical systems are more attractive from the aspect that they require smaller apertures than millimeter wave systems for the same data rate. The diameter of the telescope for an optical communications system is often an order of magnitude less than the diameter of the antenna that is needed for a comparable millimeter wave system.

These studies have usually indicated that the preferred system is a function of data rate, with millimeter wave systems being the choice for lower data rates, and optical systems being the choice for higher data rates. The "crossover" frequency at which both types of systems are equally attractive has been found to be anywhere between 50 Mbps and 2 Gbps.

As it is currently believed that both millimeter wave and optical systems will share the burden of high data rate transfer in the future, the Goddard Space Flight Center has been active in developing technology for both systems. Many of the components, such as multiplexers, demultiplexers, AGC amplifiers, and real time error performance measurement systems are applicable to both millimeter wave and optical systems.

#### TECHNOLOGY TO IMPROVE MILLIMETER WAVE COMMUNICATIONS

A typical millimeter wave communication system is illustrated in Figure 3. The high rate data modulates the millimeter wave carrier, and is then amplified before being delivered to the (transmitting) antenna system. The receiving terminal would most likely employ a low noise amplifier to improve receiver sensitivity, down convert the millimeter wave signal to an intermediate frequency (IF), and then demodulate it.

Goddard Space Flight Center's Microwave Technology Branch has been active in developing millimeter wave technology for several years. The frequency band of 59 to 64 GHz has been selected for intensive development in that it makes 5 GHz of bandwidth available, and that this

band has been allocated exclusively for space-to-space communications. The 59 to 64 GHz band is also lower in frequency than some of the other millimeter wave intersatellite bands, which makes component development somewhat easier.

#### Modulator/Exciters

The modulator/exciter generates the millimeter wave (60 GHz) signal to be transmitted, and superimposes the high data rate digital signal upon it. The modulation technique that has been selected for the development effort is quaternary phase shift keying (QPSK). QPSK was selected as it represents a compromise between efficiency (the number of bits per second that can be transmitted per unit bandwidth), and the technical complexity of the hardware. It is also used by NASA in many space communication applications.

As it was previously very difficult to directly modulate a 60 GHz carrier, initial efforts consisted of generating a high quality 15 GHz signal, modulating it, and then frequency multiplying the modulated signal by 4. This was a satisfactory approach, however, a large amount of power at 15 GHz was required due to the poor efficiency of the times 4 multiplier. This technique had one other disadvantage. The effects of any non-linearities in the modulation process at 15 GHz were multiplied by four in the frequency modulator.

With the advent of high performance, high speed schottky switching diodes, an effort was initiated to develop a modulator/exciter that would directly modulate a 60 GHz carrier. A Gunn oscillator, integrated on a quartz substrate with a coupler and a harmonic mixer, was phase locked to a 15 GHz source, which was then phase locked to a 2.14 GHz high stability source. The modulator was constructed using microwave integrated circuit techniques on a sapphire substrate. The resulting component was a modulator/exciter, with less than two degrees of non-linearity, capable of handling 2 Gbps of data on each channel, for a total of 4 Gbps.

#### Solid State Power Amplifiers

Early Goddard efforts concentrated on the silicon IMPATT (Impact Ionization Transit Time) device, and circuits that would combine the output power from multiple IMPATTs to realize higher output power. The silicon IMPATT was chosen as there was extensive reliability data on these devices, and mean device lifetime was a direct function of the operating temperature of the device.

A breadboard amplifier was developed that combined four IMPATTs in the output stage, and delivered 4 watts across a 2.5 GHz band centered at 60 GHz, with an input power of 4 milliwatts. This represented the first time ever that high power double drift IMPATT diodes were successfully used in a broadband stable amplifier. A second effort was undertaken to

combine 16 "off-the-shelf" IMPATTs using a novel radial line combining technique, to deliver 10 watts of power. Although 60 GHz combining efficiencies of over 90% were demonstrated, the completed amplifier exhibited a DC to RF efficiency of slightly less than 6 percent.

An effort is underway to develop a solid state amplifier with at least 10% DC to RF efficiency at the 10 watt level. To accomplish this, the output from eight Gallium Arsenide IMPATTs are being combined in a special low loss waveguide combining structure. Since the initial silicon IMPATT activities, much has been learned regarding the reliability associated with high power Gallium Arsenide devices, and Gallium Arsenide IMPATTs now deliver more power per device, at higher DC to RF efficiencies, than their silicon counterparts.

#### Beam Waveguide Transmission Systems

Connecting a transmitter/receiver to a millimeter wave antenna system is a formidable problem. Coaxial cable, useful at RF and lower microwave frequencies, cannot be considered for millimeter waves, due to the high loss. Waveguide, which is much less lossy than coax, is also impractical for long transmission lines. The standard waveguide which is used at 60 GHz has a loss of 0.07 dB per inch. This means that half of the received or transmitted signal will be lost for every 3.6 feet that the signal travels through such a waveguide.

To permit the transmitter/receiver to be located up to several meters from the antenna feed, beam waveguide technology is being investigated as an alternative to conventional waveguide and rotary joints. Although final measurements have not been made, an engineering model beam waveguide transmission system for 60 GHz has been fabricated which consists of mirrors, a motorized system that can slew on two axes, and a six port network for interfacing with the antenna horn.

#### Low Noise Receiver Technology

Before the advent of low noise amplifiers that operate at 60 GHz, it was envisioned that a low noise mixer front end receiver would interface directly with the antenna system. To realize small size, light weight, high performance receiving subsystems, an integrated receiver approach was taken. A low noise crossbar stripline balanced mixer was integrated with a phase locked Gunn local oscillator and a harmonic mixer on a single quartz substrate. A wideband low noise IF amplifier was fabricated on a separate substrate, and the necessary phase lock electronics and frequency source was fabricated in a separate miniature package. This technology still represents the basic configuration that will be used to down convert a 60 GHz signal to a lower IF for demodulation.

One of the first technologies investigated to improve the sensitivity of the mixer front end receiver was the

Superconductor-Insulator-Superconductor (SIS). This device, when used as a mixer, offers potential conversion gain, requires extremely low amounts of local oscillator power, and offers potential sensitivities near the fundamental quantum limit. Unfortunately, the device only operates at extremely low physical temperatures. The unavailability of long lifetime closed cycle refrigerators for spacecraft use, plus the availability of a new semiconductor device, the High Electron Mobility Transistor, has shifted attention away from the SIS for spacecraft applications, at least for the future.

Three terminal devices were also investigated as a means of improving receiver sensitivity. An effort was undertaken to improve conventional GaAs MESFETs (Gallium Arsenide Metal Semiconductor Field Effect Transistor) by shortening the gate length, decreasing the gate resistance, and decreasing the gate capacitance. While good results were obtained at 60 GHz, a new device became available which exhibited even better performance in terms of sensitivity. This device, the High Electron Mobility Transistor (HEMT), has become the best device for achieving low noise gain at frequencies up through about 90 GHz. Considerable effort was spent in developing state-of-the-art devices and circuits. Single, two, and three stage amplifiers were fabricated and tested. The resulting amplifiers are capable of improving the sensitivity of the previously developed mixer front end receiver by a factor of about five.

#### TECHNOLOGY TO IMPROVE OPTICAL AND MILLIMETER WAVE COMMUNICATIONS

Goddard's Microwave Technology Branch has also been developing high speed electronics that are applicable to both optical and microwave/millimeter wave communication systems. These electronics are used for multiplexing data from several sources, encoding the data for transmission, adjusting the received signal levels, recovering the data and clock signals, and providing a means of measuring the performance of the system in real time.

Figure 5 represents the electronics that are associated with either an optical or a microwave/millimeter wave system. The multiplexer combines digital data from several sources. The multiplexed signal is then encoded. Optical systems are being designed which use a pulse position modulation format. This is critical to optical systems to limit the average power from the laser transmitter, while keeping the ratio of peak to average power as high as possible. An automatic gain control (AGC) amplifier is used to adjust the received signal to deliver a constant output to the bit synchronizer. Such adjustments are necessary due to fluctuations in input signal from the pointing, acquisition, and tracking system associated with the optical telescope, or the millimeter wave antenna. The bit synchronizer recovers the transmitted data and clock. The data can then be demultiplexed to reconstruct the original sources. A pseudo random code generator and bit error detector provides



a means of measuring the real time performance of the system.

Communication electronics components that perform all of the above functions have been designed for an optical system that operates at 50 Mbps and an optical system that operates at 650 Mbps. Components have been fabricated and tested for a breadboard system that operates at 50 Mbps. These components include a binary pulse position modulation encoder, an AGC amplifier, data detection filters, and a bit synchronizer. Development in this area has used conventional emitter coupled logic and complementary metal oxide silicon (CMOS) logic devices for the components that operate at 50 Mbps. To operate at 650 Mbps, high speed Gallium Arsenide devices are used with fabrication implemented on multilayer printed circuit boards.

#### POTENTIAL COMMERCIAL APPLICATIONS

All of the technology previously described has commercial applications. While the components developed for space-to-space communications at 60 GHz may not have specific commercial applications, it is very possible to modify them slightly for operation at different millimeter wave frequencies. This would enable them to be used for extremely wide band, high data rate transmissions over short to moderate (line of sight) distances. The local networking of several high speed computer centers is one possible application.

To serve the remote user in an isolated area where conventional hard wired or fiber optic facilities would be prohibited, microwave and millimeter wave systems would enable the user to receive large quantities of high rate data in short periods of time. Such a system could also be interactive, or bi-directional, allowing the remote user to transmit either low data rate requests or large volumes of data at high speed to another location.

The high speed communication electronics work is also directly applicable to commercial users. Much of this technology is just as applicable to fiber optic transmission as it is for transmission through the atmosphere. High data rate terminals could be developed that would allow local transmission of high data rates between multiple nodes. Such systems could be built with adequate margin to operate over short distances in the presence of moderate precipitation. If these margins are not incorporated into the system, use of the system could be restricted to periods when precipitation or fog are not intense enough to preclude accurate transmission of data.

The use of high data rate communication systems will become increasingly important in the future. Sophisticated communication devices of the future will require it. Increasing costs of transportation will make it economically desirable. Uses of these systems will only be limited by the imagination of the system designers of the future.

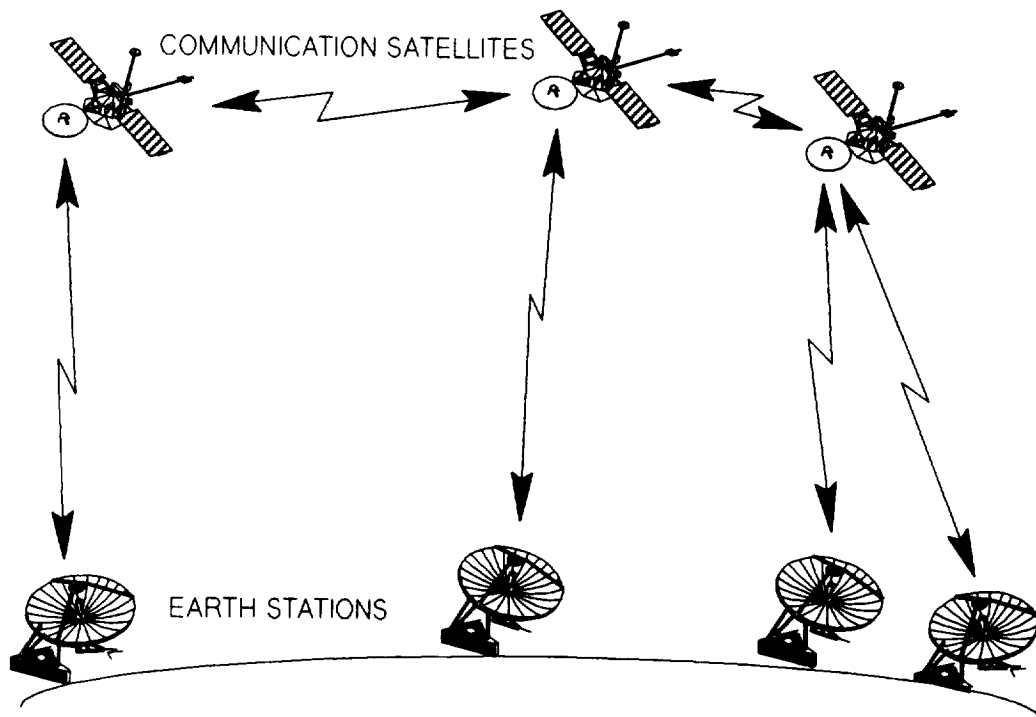


FIGURE 1. COMMERCIAL SATELLITE BASED SYSTEM

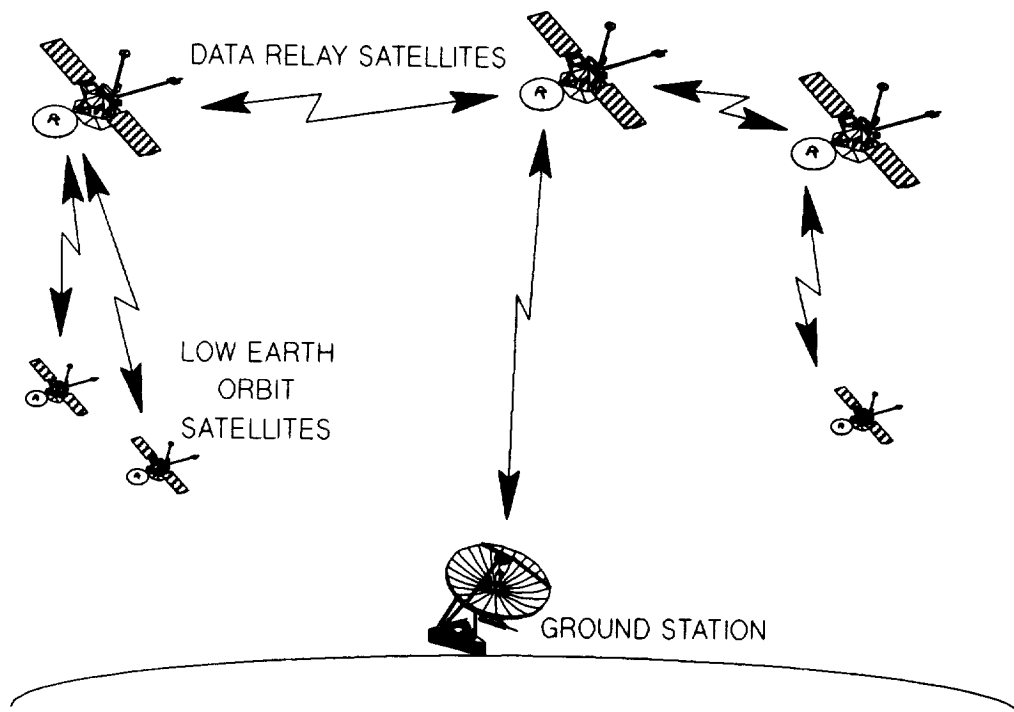


FIGURE 2. NASA SATELLITE BASED SYSTEM

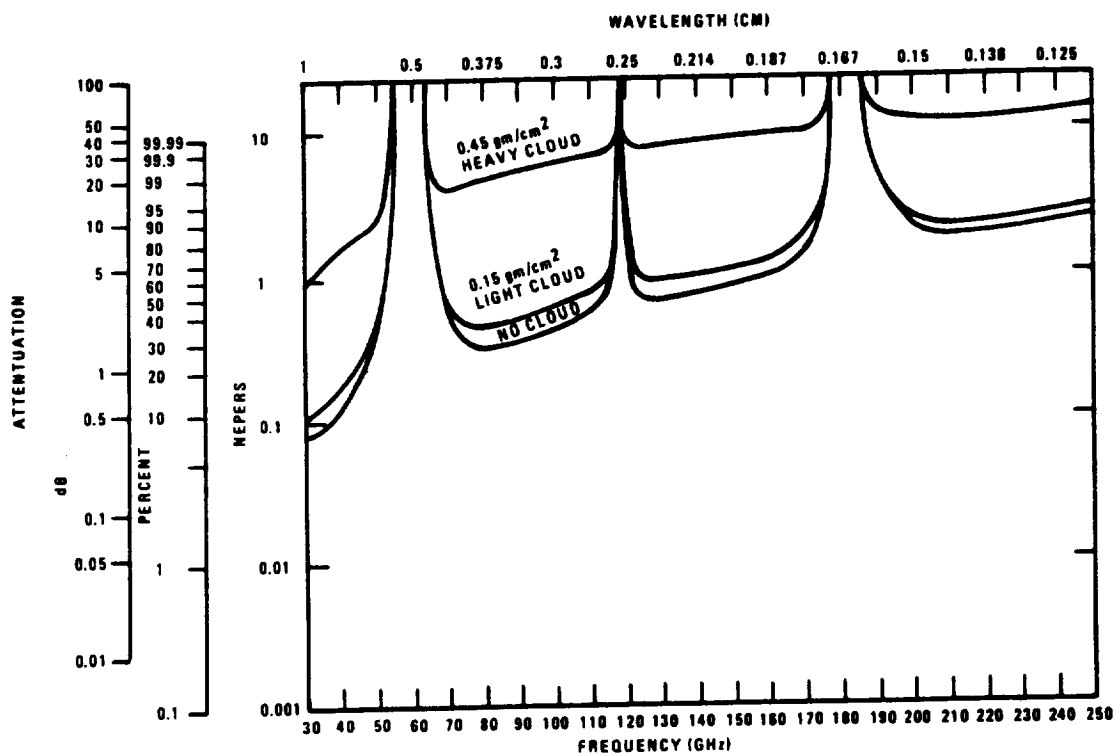


FIGURE 3. ATMOSPHERIC ATTENUATION

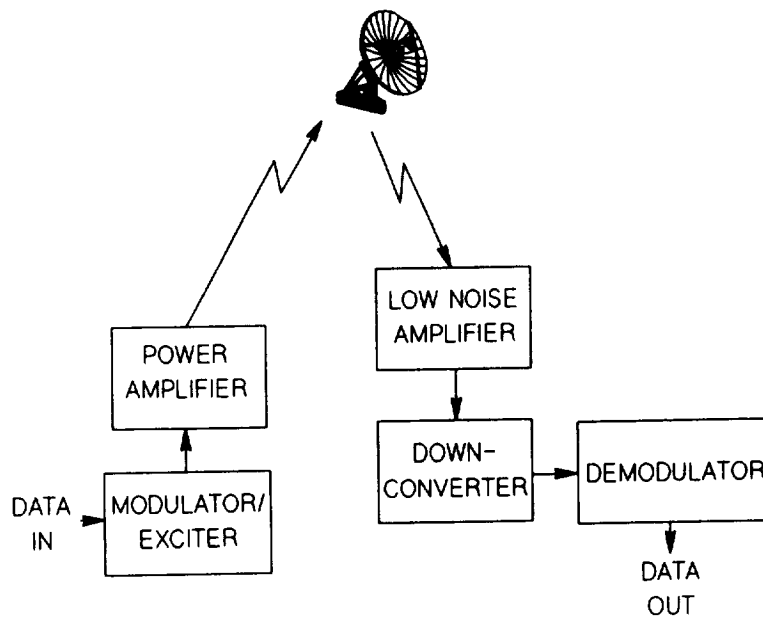


FIGURE 4. MILLIMETER WAVE SYSTEM

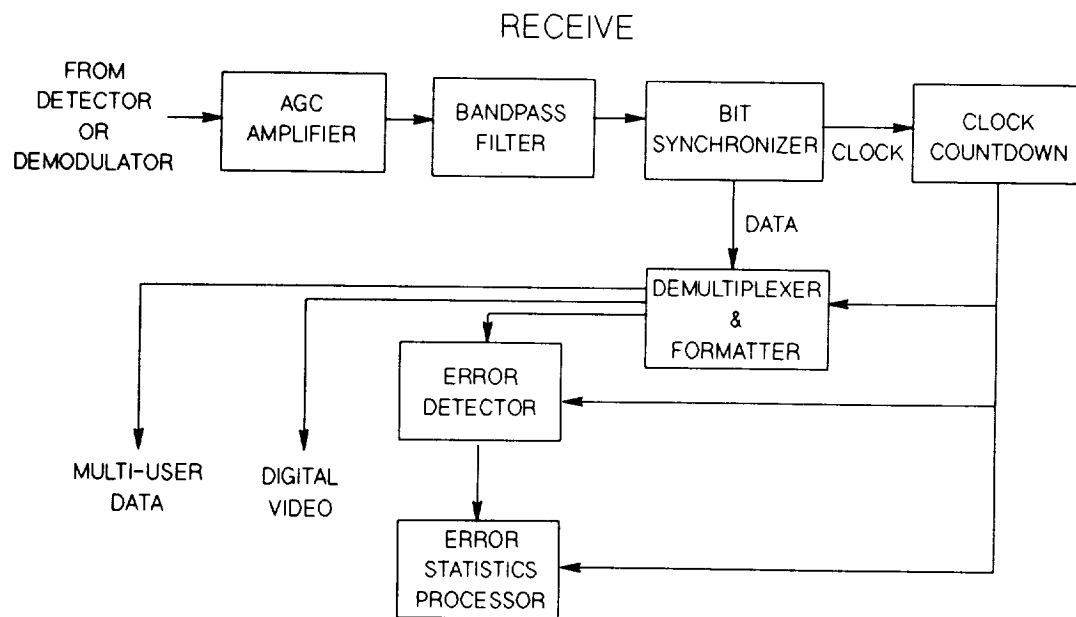
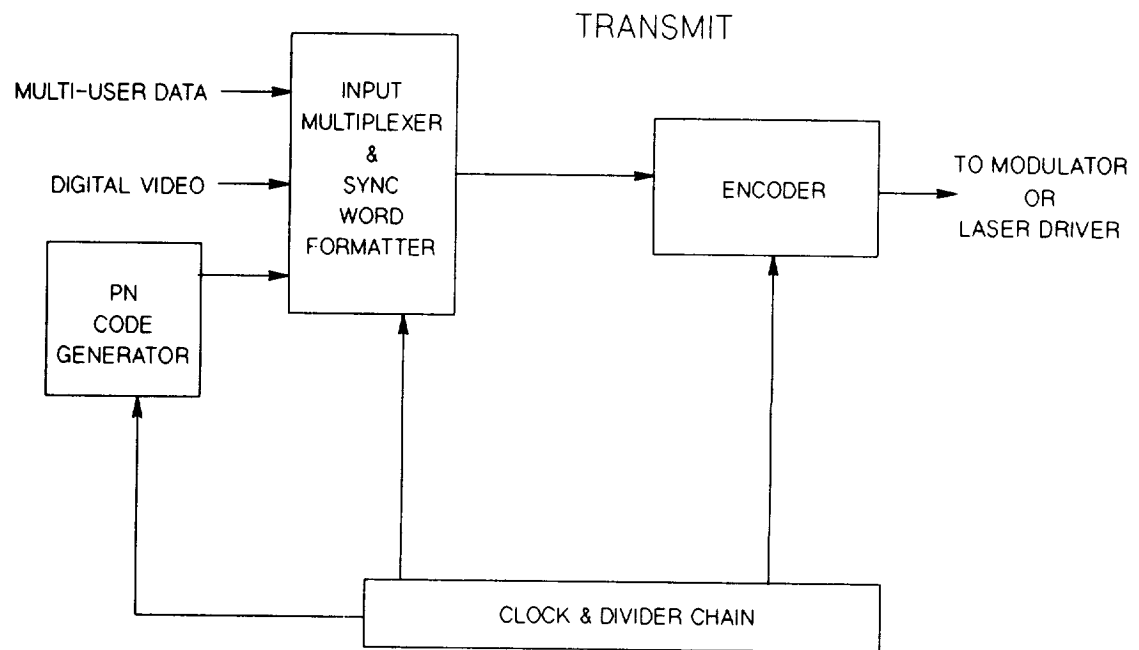


FIGURE 5. TYPICAL COMMUNICATION ELECTRONICS

## DISTRIBUTED ACCESS VIEW INTEGRATED DATABASE (DAVID) SYSTEM

Barry E. Jacobs  
Code 934  
Goddard Space Flight Center  
Greenbelt, MD 20771

### Introduction

Due to the diversity of computers, operating systems, management systems, network protocols, etc., NASA space scientists have to learn many different access methods in order to obtain data. For example, in NASA's Space Astrophysics Program, Astrophysics observatories such as the International Ultraviolet Explorer (IUE), Hubble Space Telescope, Gamma Ray Observatory (GRO) etc, generally have one or more data centers for the analysis and distribution of data. The heterogeneity of its data centers makes Astrophysics multi-mission research almost impossible.

The Distributed Access View Integrated Database (DAVID) System, which is the subject of this paper, is a solution to the problem. See Figure 1. The DAVID System has been adopted by NASA's Astrophysics Division for their Astrophysics Data System and represents one of the most extensive national testbeds of heterogeneous distributed computing systems. Although the DAVID System was originally built to uniformly access databases, it is currently being generalized to other object types such as spreadsheets, images, manuscripts, graphics, etc. The DAVID software is in the process of being rebuilt by the private sector for adoption by other NASA disciplines.

The DAVID approach to the heterogeneous distributed systems problem is at four levels. See Figure 2. At the lowest level, we develop universal object type management systems to provide uniform access to heterogeneous database, images, spreadsheet, manuscript, etc. management systems. At the second level, we develop "book" and "kit" management systems to provide uniform access to aggregate sets of data objects. At the third level, we develop "libraries" to provide uniform access to a local area networks of computers containing "books", "kits", and other data objects. At the fourth level, we develop consortiums of libraries to provide access to sets of libraries.

An outline of the paper is as follows. In Section 1, we outline the heterogeneous components of the Astrophysics program. In Section 2, we describe the Library and Library Consortium levels of the DAVID approach. In Section 3, we discuss the Books and Kits level. In Section 4, we describe the Universal Object Type Management System level. In Section 5, we explain the relation of the DAVID project with the Small Business Innovative Research (SBIR) Program. We conclude the paper in Section 6, with a summary.

### The Heterogeneous Astrophysics Program

NASA's Astrophysics program provides a good example of the system heterogeneity problem. We briefly outline some of the heterogeneity.

The Infrared Processing and Analysis Center (IPAC) at California Institute of Technology manages data from infrared sensors such as the Infrared Satellite (IRAS). IPAC has Cyber and Sun computers, Cyber and Unix Operating Systems, IM/DM database management system, and Internet and Dec Net communications. The data includes IRAS observation logs (databases), IRAS products (databases), IRAS imagery (images), IRAS spectra (spreadsheets), IRAS documentation (manuscripts), and IRAS software.

The Smithsonian Astronomical Observatory (SAO) at Harvard University manages data from the high-energy sensors such as the Einstein or High-Energy Astronomical Observatory (HEAO). SAO has VAX, Sun and Cyber computers, VAX/VMS and Unix Operating Systems, Ingres database management system, and Internet and Dec Net communications. The data includes Einstein observation logs (databases), Einstein products (databases), Einstein imagery (images), Einstein spectra (spreadsheets), Einstein documentation (manuscripts), and Einstein software.

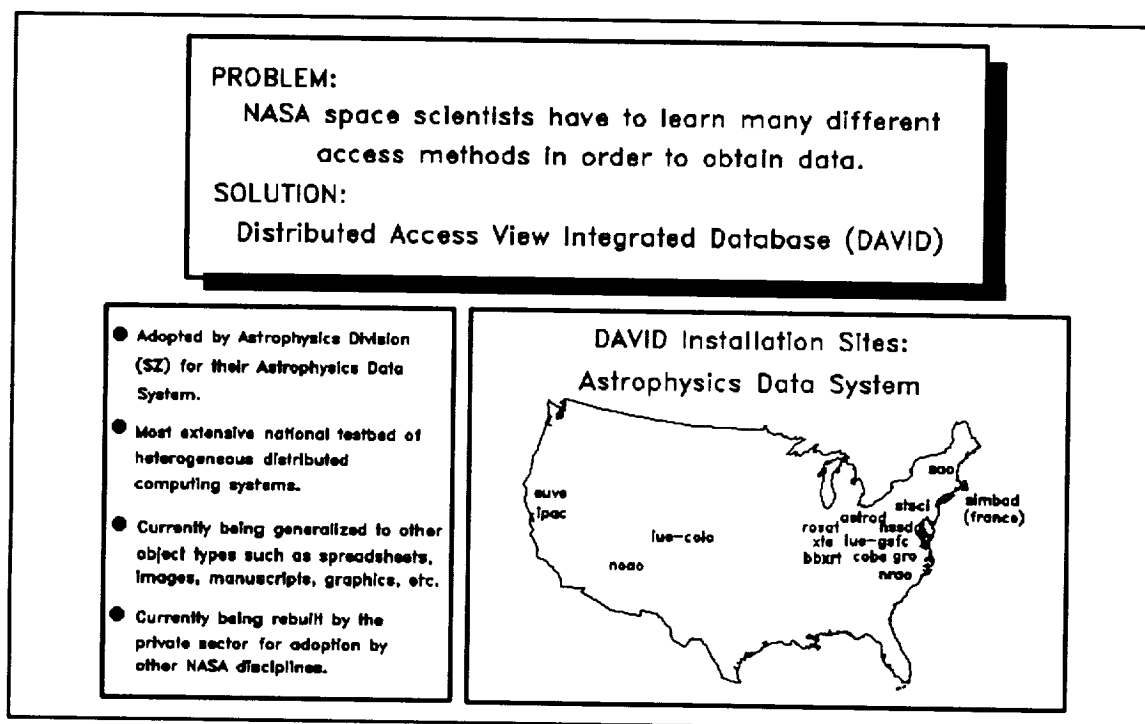


Figure 1.

DAVID Approach To The Heterogeneous Distributed System Problem		
Level	Model	Examples
4	<b>Consortiums Of Libraries:</b> Aggregate Set of Related Libraries	ADS, High-Energy, Gamma Ray
3	<b>Libraries:</b> Collections of Holdings	NSSDC, IUE RDAF, IPAC, SAO, StSCI
2	<b>Book &amp; Kit Objects:</b> Aggregate Set of Related Objects	Library Catalog Book, IUE Observation Book, IUE Software Tools Book, Astrophysics Research Kit
1	<b>Universal Object Type Management Systems:</b> Provides Uniform Access to Heterogeneous Objects	Universal Database Management System, Universal Image Management System, Universal Manuscript Management System

Figure 2.

The International Ultraviolet Explorer (IUE) Regional Data Analysis Facilities (RDAF) at the Goddard Space Flight Center and at the University of Colorado manage data from the International Ultraviolet Explorer mission. The RDAFs have VAX, Sun and IBM computers, VAX/VMS, Unix, and IBM Operating Systems, Ingres, IBM database management system, and Internet and Dec Net communications. The data includes IUE observation logs (databases), IUE products (databases), IUE spectra (spreadsheets), IUE documentation (manuscripts), and IUE software.

The Space Telescope Science Institute (STScI) at the Johns Hopkins University manages data from the Hubble Space Telescope (HST). The STScI has VAX and Sun and Cyber computers, VAX/VMS and Unix Operating Systems, Sybase and Britten-Lee database management systems, and Internet and Dec Net communications. The data includes HST observation logs (databases), HST products (databases), HST spectra (spreadsheets), HST documentation (manuscripts), and HST software.

The National Space Science Data Center (NSSDC) at the Goddard Space Flight Center manages data from a number of astrophysics and non-astrophysics missions. The NSSDC has VAX, Sun, IBM, and Cyber computers, VAX/VMS, IBM and Unix Operating Systems, Ingres, Oracle, Sybase and Britten-Lee database management systems, and Internet and Dec Net communications. The data includes multi-mission observation logs (databases), data products (databases), spectra (spreadsheets), documentation (manuscripts), and software.

### **Libraries and Library Consortia**

The DAVID approach, models a local area network as a library and a set of local area networks as a consortium of libraries. See Figure 3.

Each local area network of computers is modeled as a **library** in which each of the machines is a "room". One particular room, the "main room", serves as a gateway into the library. When one logs on to the main room and enters the command "dlib", one gets the Library Directory menu which leads the user to the Main Room (described below), Tutorials, Reading Rooms (computers for library holdings), Personal Library Rooms (computers for user holdings), and Management Rooms Menus (computers for staff holdings).

When the Main Room option is chosen, the user sees the Main Room Menu. This consists of submenus for Administrative Desk (library cards, bulletin board, mail, phone, suggestions, etc.), Reference Desk (names, phone and forms to reference librarians), Library Catalogue (database of local library holdings), Circulation Desk (forms for borrowing off-line holdings), Reproduction (forms for reproducing off-line holdings and software for reproducing on-line holdings), Union Library Desk (access to union catalogues and facilities for consortia), and Remote Libraries Desk (access to remote libraries).

A local library may be a member of one or more **consortia of libraries**. A library consortium is a organization of participating libraries which share each other's resources. One of the functions of a library consortium is to provide a **union catalogue** of all of the holdings over all of the member libraries. A user can access a union catalogue of a consortium and determine which member has the holding. He can then ask his local library to help him obtain access to it.

For example, in the Astrophysics example, the library representing the Center For Astrophysics (CFA) could belong to several consortia, for example, "X-Ray", "High-Energy", "Astrophysics", and "Multi-Disciplinary". Then if a user at CFA wants a holding not at CFA he can consult one or more of the union catalogues of the different consortia. He would then locate the holding and ask his local librarian to obtain access to the holding.

### **Books and Kits**

The DAVID approach, models an aggregate of related data as a book and a kit.

A **book** is an aggregate data structure which is divided into logical units called **chapters** and physical units called **components**. An example of a book is given in Figure 4. for the Book of IUE Observations. Here each chapter represents one of the 80,000 observations taken by IUE. The IUE book has as components- a title page (manuscript), a preface (manuscript), a table of contents (database), chapter components including- observation descriptions (manuscript), proposals (manuscripts), literature references (database), raw image (images), extended-line-by-line spectra (spreadsheets), observation scripts

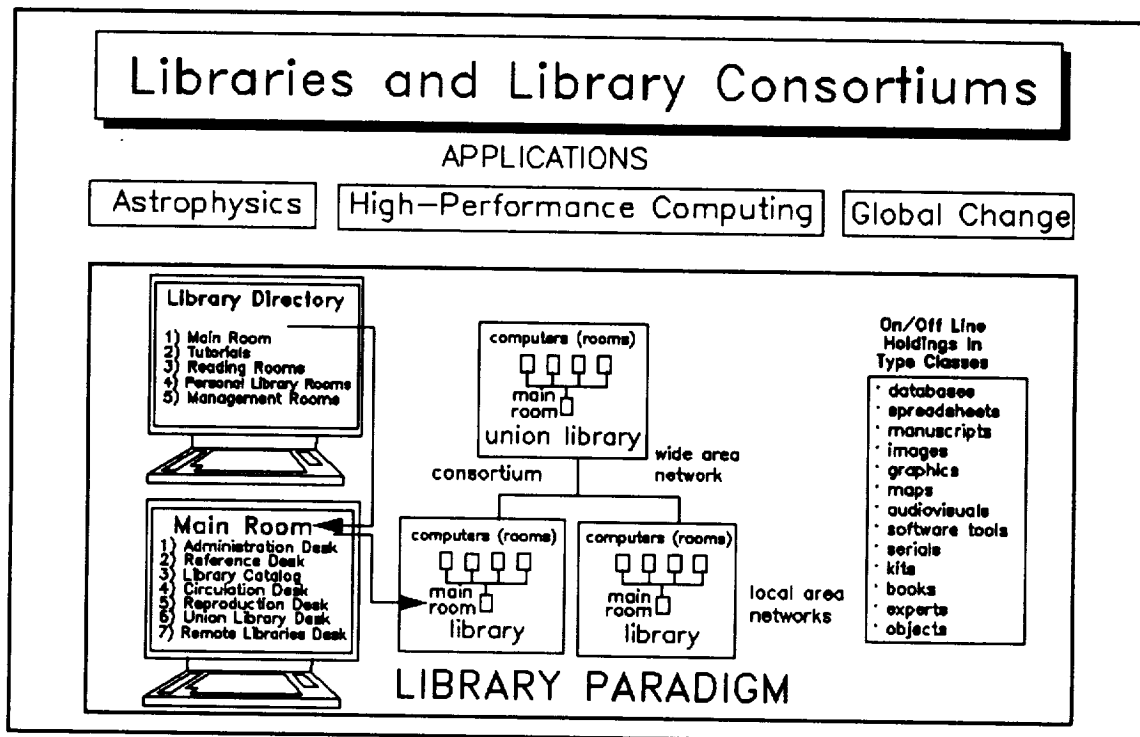


Figure 3.

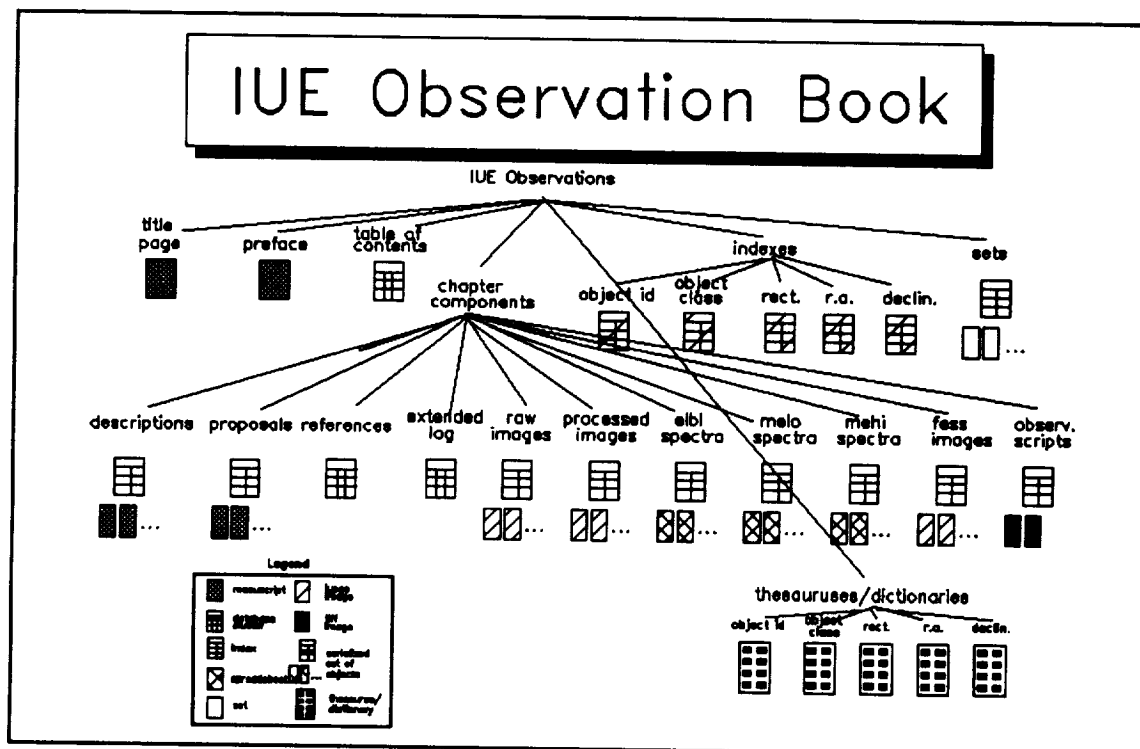


Figure 4.



(images), indexes including object id, object class, rectangle. In addition, each of the index components has a thesaurus/dictionary database. Another example of a book is the IUE Software Tool Book where each chapter represent one of 900 software programs written expressly for IUE data. A third example is the Local Library Catalogue Book where each chapter represents one of the holdings at the local area network.

A kit is an aggregate data structure which is just divided into physical units called components. An example of a book is given in Figure 5. for the IUE Software Tools kit for managing IUE software. The IUE Software Tools kit has as components- a title page (manuscript), a preface (manuscript), requirements document (manuscript), users guide (manuscripts), high-level design document (manuscript), system administrators guide (manuscript), users comments (manuscript), reference manual (book), low-level internals (book), installation (book), source code (book), object code (book), testing plan (book), and software tool (book). Another example of a kit is an Astrophysics Research Kit which represents an Research problem in Space Astrophysics and whose components are subkits each representing a sub-problem of the original problem.

### **Universal Object Type Management Systems**

The DAVID approach, make use of universal object type management systems.

A **Universal Object Type Management system** allows one to retrieve, read and operate on object of a particular object type independent of the subtype and internal format. Figure 6. contains examples of object types for which universal object type management systems are important. For example, a universal database management system allows one to operate on databases independent of the type (e.g., relational, hierarchical, and network) and internal format (e.g., Ingres, IM/DM, Sybase, IMS, etc.). A universal manuscript management system allows one to operate on manuscripts independent of the type (e.g., structured, unstructured) and internal format (e.g., ASCII, Tex, Troff, Word Perfect, etc.). A universal image management system allows one to operate on images independent of type (e.g., level 0, level 1) and internal format (e.g., TIFF, FITS, IRAF, etc.).

### **DAVID And the SBIR Program**

The DAVID project has made strong use of the Small Business Innovative Research (SBIR) Program. A number of successful SBIR projects are described here.

**Heterogeneous Network Access** by Digital Analysis Corporation of Reston, Virginia. The purpose of this project was the development of a uniform layer that can be put on top of heterogeneous network alternatives such as TCP/IP, DEC Net, Bit Net, etc. Each subnet provides basic network services and hops between subnets are handled by servers at the gateways.

**Heterogeneous Interface Building** by K. Wanderman and Associates of Staten Island, New York. The purpose of this project is the development a software that will facilitate development of interfaces of database management systems and arbitrary files into the NASA developed Universal Database Management System. The software will be playing a central role in the establishment of major repository of Astronomical Catalogues at the NSSDC.

**Universal Index Management Systems** by Advanced Communication Technology Incorporated of Silver Spring, Maryland. The purpose of this project is the development of a universal index management system. The software will allow users and system builders to construct indexes of different types (one, two, three dimensional, etc) and different formats (B-trees, R-trees, hash, etc.). The software will be playing an important role in speeding up access of data in the Astrophysics Data System.

**Universal Software Tool Management Systems** by HSA Incorporated of Sugar Land, Texas. The purpose of this project is the development of a universal software tools management system for managing software over distributed systems. Each software tool is represented as a kit in a similar way that the IUE Software Tools Kit above was described. The software will be playing an important role in the management of software in the Astrophysics Data System.

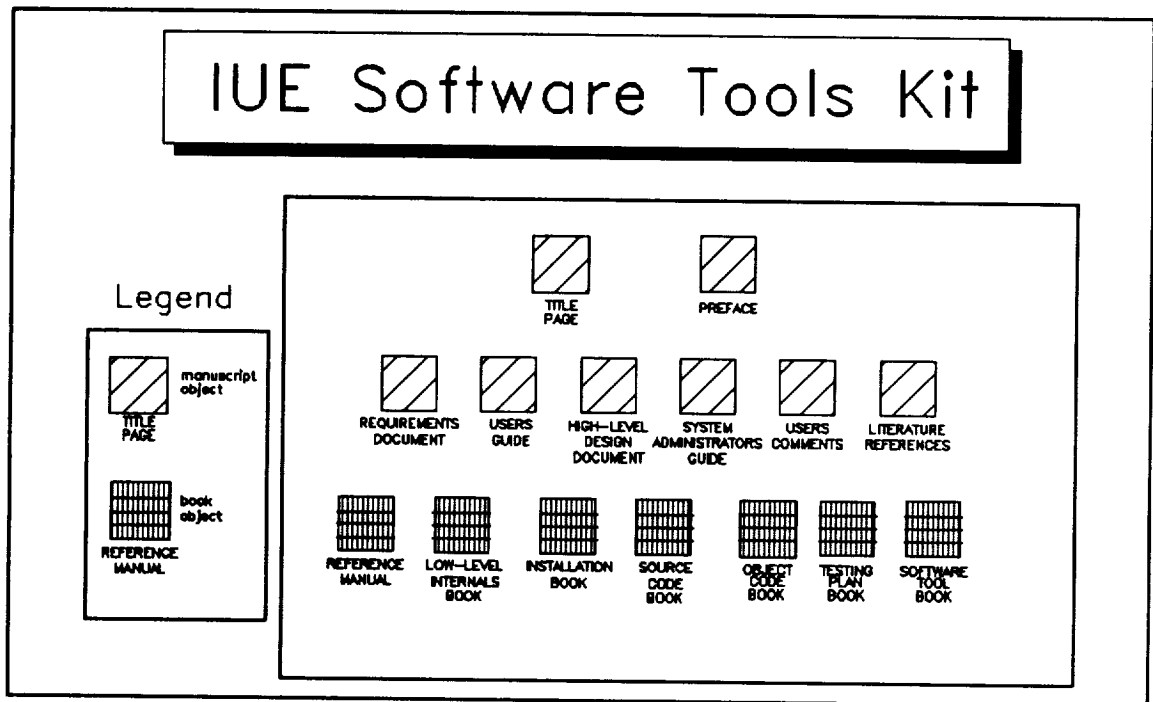


Figure 5.

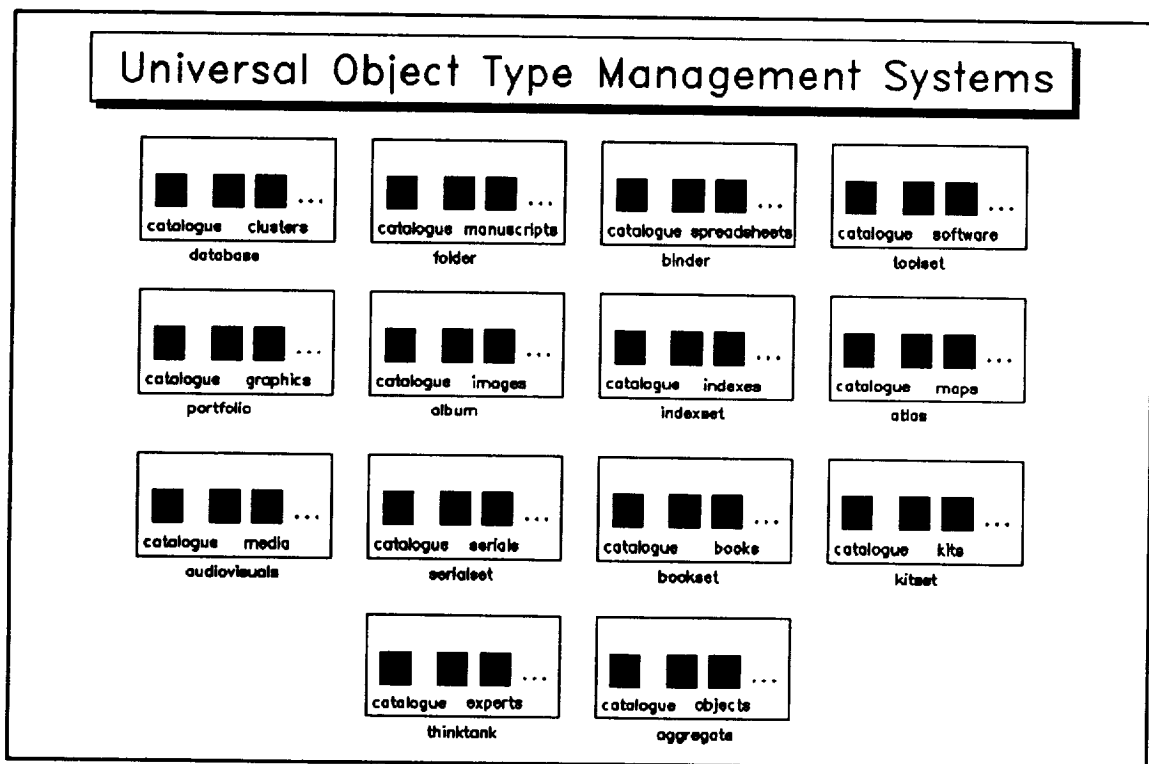


Figure 6.

**Universal Object Type Management Systems** by REI Associates of McLean, Virginia. The purpose of this project is the development of a management system for building other management systems. Each management system is represented as a kit which contains retrieval software tools and request server tools. The software will be playing an important role in the development of other Universal Object Type Management Systems.

**Universal Book Management Systems** by Advanced Applications Corporation of Potomac, Maryland. The purpose of this project is the development of a universal books management system for managing book objects. Each book is represented in a similar way that the IUE Observation Book above was described. The software will be playing an important role in the management of data in the Astrophysics Data System.

### **Summary**

The Distributed Access View Integrated Database (DAVID) System, which has been adopted by the Astrophysics Division for their Astrophysics Data System, is a solution to the system heterogeneity problem. In Section 1, we outlined the heterogeneous components of the Astrophysics problem. In Section 2, we described the Library and Library Consortium levels of the DAVID approach. In Section 3, we discussed the Books and Kits level. In Section 4, we described the Universal Object Type Management System level. In Section 5, we explained the relation of the DAVID project with the Small Business Innovative Research (SBIR) Program.

## Three-Dimensional Perspective Visualization

**Kevin Hussey**  
**Supervisor of the Visualization and**  
**Earth Science Applications Group**  
**Jet Propulsion Laboratory.**

### Abstract

It has been demonstrated that image processing/computer graphic techniques can provide an effective means of physiographic analysis of remotely sensed regions through the use of three-dimensional perspective rendering. This talk will explain the methods used to simulate and animate three-dimensional surfaces from 2-dimensional imagery and digital elevation models. A brief historic look at JPL's efforts in this field and several examples of animations, illustrating the evolution of these techniques from 1985, will be shown. JPL's current research in this area will also be discussed along with examples of technology transfer and potential commercial application. The software is part of the VICAR image processing system which was developed at the Multimission Image Processing Laboratory of JPL.

### Narrative Talk Outline

- I. Historic Perspective from 1985: How and Why we began work in 3-dimensional perspective rendering (3DPR).
  - A. JPL's Atmospheric Parameter Mapping Task were producing several 2-dimensional time series animations of several atmospheric parameters derived from satellite data.
  - B. A primary visualization challenge was to simultaneously display multiple parameters like percent cloud cover and cloud top barometric pressure.
  - C. The Animation "Topo Follies" by JPL Scientist Dr. Mike Kobrick was produced in 1985. This was the first 3DPR animation utilizing Kobrick's original ray casting program. It simultaneously displays LANDSAT multispectral image data with a digital elevation model used as the third dimension "Topo Follies" depicts a short flight around and through Death Valley California.
  - D. The above technique was imported into the VICAR (Video Image Communication and Retrieval) image processing software system, developed by JPL to support planetary exploration, and applied to cloud cover and cloud top barometric pressure data. The inverse of barometric pressure was used as the third dimension. Several 3DPR animations of cloud cover and altitude have been produced.
- II. Improvement of 3DPR capabilities: Improvements have come in 3 categories. The first to be dealt with was ease of use (A thru D below), next came speed of rendering (E below) and finally the quality of the rendering (F & G below).
  - A. Original FORTRAN code from Kobrick required fixed input and output image sizes along with pre-computed "flight" parameter tables.
  - B. The VICAR version of render was re-written in C and allows variable sized inputs and outputs. (Largest inputs and outputs used to date are: 20,000 lines by 14,000 samples input, and an output image size of 3100 lines by 4200 samples). There are no set size limits.
  - C. Interactive "Flight" parameter selection software written: The program 3DI

(three-dimensional interactive) is used to select the viewing parameters. 3DI graphically represents the parameters on a workstation display monitor and allows the user to change them interactively. Once satisfied with the changes, the user can preview the resulting 3DPR image as calculated from the actual input images to be used for the animation. After previewing, the parameters and/or previews may be saved for later application. 3DI is not used to generate the movie frames because it works on a scaled down version of the original input image and elevation data. This is done to greatly reduce the amount of processing speed needed to generate the previews.

- D. "L. A.: The Movie" was the result of the first application of 3DI and VICAR render. It's a 2 minute and 6 second flight simulation over the Los Angeles area using LANDSAT Thematic Mapper data and digital terrain elevation data.
- E. The "Pyramid" rendering algorithm: A major speed up. The new algorithm uses a multiple resolution input image pyramid to improve rendering speed and quality. One scene, rendered from a LANDSAT mosaic of California, took 4 hours and 44 minutes to render before the pyramid algorithm was implemented; the same scene took only one minute and 30 seconds using an eight resolution level input data pyramid. This represents a speed-up of 189 times. In addition to the speed up, the animation quality was greatly improved by reducing high frequency aliasing artifacts (sparkle).
- F. Multiple Rays per pixel: For each pixel in the output image, several rays (instead of just one) are extended from the "camera" through the output pixel and lengthened until they intersect the surface defined by the elevation data. The rays, in effect, form a cone that intersects with the surface. The area inside the "cone" on the input image are averaged to produce the value of the output pixel. This technique also greatly reduces the amount of high frequency aliasing artifacts (sparkle) which results in greater visual clarity.
- G. Field Rendering: Doubling the number of computed fields per second nearly eliminates jitter. A standard NTSC television signal is comprised of 30 frames per second. Each frame is comprised of 2 interlaced fields. By computing and recording the 60 interlaced fields per second instead of the 30 frames you can virtually eliminate the appearance of objects jittering on the screen. This improvement has been incorporated into the 3DPR software.

- IV. Current Research: Along with continuing to improve the ease of use, speed and quality of the 3DPR techniques, we are adding additional capabilities to our 3DI program. These capabilities include the ability to "interactively fly" over very large data bases under joy stick control and to interactively perform basic analytical functions on user selected areas of the 3DPR images being displayed. (The level of interactivity will be, in part, a function on the workstation being used.) Volumetric visualization of oceanographic data is also currently being pursued in the 3DI context. This work is being accomplished under UNIX in an X-windows environment.
- V. Initial Technology Transfer: JPL to Syracuse Research Corporation (SRC). A task to transfer the 3DPR capability described above to private industry has begun. SRC is funding a portion of the current visualization research taking place at JPL in return for a license to utilize the prototype technology developed for commercial applications. The benefits to JPL are two fold. First, it provides additional funds for research and second, SRC becomes an active participant in the research, through their own independent development, and exhaustive testing of the new capabilities.
- VI. Conclusions: We've only just begun. Even though we've been working in the area of 3DPR since 1986 (at less than a one person per year effort) we only recently have received funding to proceed with our planned improvements at a three-person per year level for at least 1991. We believe adding interactivity, volume visualization and a built in analytical capability, will make the

three-dimensional perspective rendering techniques described here a valuable contribution to NASA's goal of providing scientists with the technology they need to understand some of the Earth's problems.

## **SESSION D - MATERIALS SCIENCE (PART 1)**

**Tuesday November 27, 1990**

- **High-Performance Polymer Development**
- **Industrial Applications Of Graphite Fluoride Fibers**
- **Unique Applications Of Fluorepoxy Materials**
- **Dual Beam Process Diamond-Like Films For Industrial Applications**
- **Plasma-Polymerized Coating For Polycarbonate: Single-Layer, Abrasion-Resistant, And Antireflective**
- **The PM200 Lubrication System**





## HIGH PERFORMANCE POLYMER DEVELOPMENT

Paul M. Hergenrother  
Materials Division  
NASA Langley Research Center

## INTRODUCTION

The term high performance as applied to polymers is generally associated with polymers that operate at high temperatures. This is somewhat misleading since there are many polymers that can be classified as high performance materials that perform well in a hostile environment where the temperature is not high. For example, polymeric separation membranes that are resistant to strong chemical action for long periods under stress at ambient temperature or composite matrices on commercial airplanes that must perform over the temperature range of -54 to 93°C for tens of thousands of hours under stress can be classified as high performance polymers. There is no standard definition for high performance polymers because the requirements and environments for different applications vary significantly. In this paper, high performance will be used to describe polymers that perform at temperatures of 177°C or higher. In addition to temperature, other factors obviously influence the performance of polymers such as thermal cycling, stress level, and environmental effects (e.g. moisture, chemical and electrical action).

In regards to high temperature polymers, the worldwide market in 1988 was estimated at 205 million pounds with a value of \$2.3B (ref. 1). This market is expected to double by the end of this decade. New applications for existing materials as well as the introduction of new polymers will contribute to this market increase. High temperature polymers are currently available in many different forms such as adhesives, coatings, composite matrices, fibers, films, foams, membranes and moldings. They are in use or are being considered for use in the following representative applications:

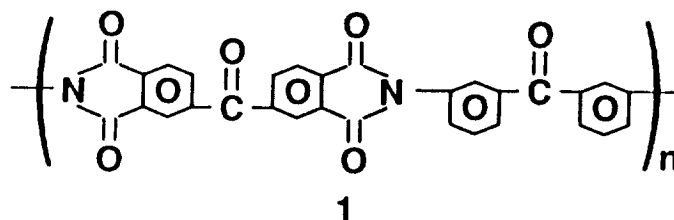
- o Electronic and microelectronic components (circuit boards, moldings, flexible cables, insulators, coatings, wire wraps, etc.).
- o Gaskets, sealants and tubing.
- o Binding systems in brake shoes, abrasive wheels and cutting discs.
- o Structural resins (adhesives, composite matrices, foams) for aircraft, space vehicles, and military items.
- o Jet engine components (fan blades, flaps, ducting, bushings, cowlings, races, etc.).
- o Nuclear reactor components (coolants, insulation, structural parts).
- o Conveyor belts for treating and drying materials.
- o Pipes for chemical processing and energy generators.
- o Fire-resistant materials (protective clothing).
- o Reinforcements (high modulus and high strength fibers and ribbons).
- o Ablators (thermal protection systems).
- o Automotive components (connecting rods, wrist pins, pistons, switches, electrical components).

- o Household items (nonstick interior and decorative exterior surfaces on cookware, irons).
- o Medical devices (prostheses, circuitry in bio-implants, dental instruments).
- o Copier machines (gears and picks).

This paper will discuss some recent developments at NASA Langley in polyimides, poly(arylene ethers) and acetylenic terminated materials. Although our work concentrates on the development of technology for aeronautical applications, primarily structural adhesives and composites, these polymers are potentially useful in other applications.

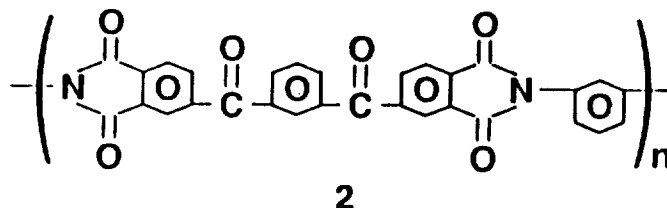
## DISCUSSION

Polyimides constitute a major class of high temperature polymers with a market value in 1988 of \$550M. Two principal uses of polyimides are films and moldings. Recent work in our laboratory on new polyimides has been directed towards the development of adhesives, composite matrices, films and moldings. A polyimide designated LARC-TPI (Langley Research Center-Thermoplastic Polyimide) as depicted in structure 1 was initially prepared in our laboratories in the mid 1970s (refs. 2-5). Since the early work on



LARC-TPI, several advances have been made. For example, the molecular weight of the polymer has been controlled and the molecules end-capped to provide a form with a lower melt viscosity (better compression and injection moldability) and better melt stability than high molecular weight unend-capped LARC-TPI. The properties of LARC-TPI are summarized in Table 1. The excellent retention of adhesive properties at 232°C after 37,000 hours at 232°C in air as well as the preliminary laminate properties are particularly noteworthy. LARC-TPI has been licensed and is available as a film, powder, and solution. This material has been successfully developed into a form for injection molding.

An isomeric form of LARC-TPI designated LARC-ITPI and shown in structure 2 has been developed.



The properties of LARC-ITPI are essentially identical to LARC-TPI. The advantage of LARC-ITPI over LARC-TPI is the use of a less toxic diamine (m-phenylenediamine) to make the polymer and the potential of lower cost. LARC-TPI is prepared using 3,3'-diaminobenzophenone, a diamine that exhibited a positive Ames test (mutagenic).

A polyimide similar in chemical structure to LARC-TPI is polyimidesulfone (PIS<sub>o</sub>2). The carbonyl group from the diamine portion of the molecule has been substituted with a sulfone group by using 3,3'-diaminodiphenyl sulfone in place of 3,3'-diaminobenzophenone. The sulfone diamine is relatively inexpensive and is non-mutagenic. The mechanical properties of PIS<sub>o</sub>2 are similar to those of LARC-TPI.

Another polyimide designated LARC-CPI (Langley Research Center-Crystalline Polyimide) as depicted in structure 3 has also received considerable attention in our laboratories. The properties are summarized in

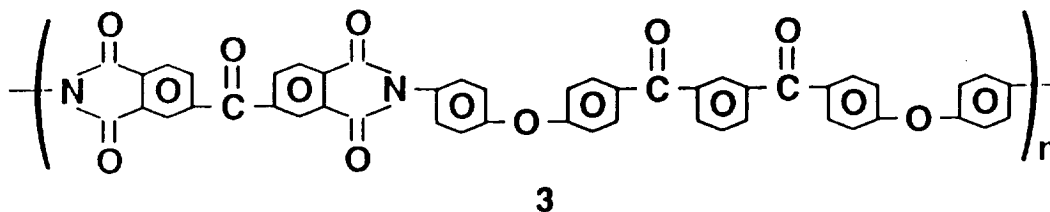
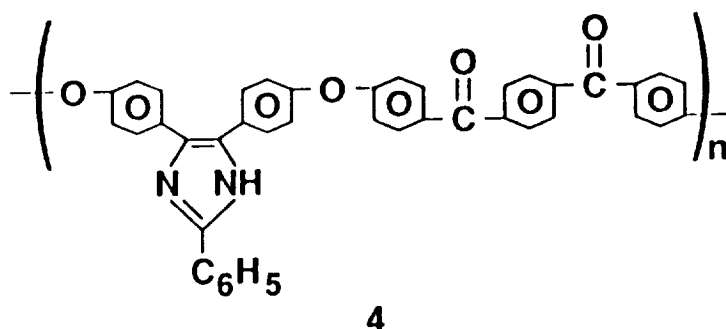


Table 2. The fracture energy as determined on compact tension specimens was extremely high due to the generation of new surface area through cracks and yielding as the crack propagated through the specimen. Films of LARC-CPI have been uniaxially oriented to give 25°C tensile strength and modulus as high as 448.1 MPa (65 Ksi) and 9.65 GPa (1,400 Ksi) respectively. The adhesive properties in Table 1 are impressive. To attain high strength at 232°C, the bonds must be annealed to induce crystallinity in LARC-CPI. Further work has involved molecular weight control and end-capping to provide a form of LARC-CPI that has better compression moldability and faster crystallization rates than high molecular weight LARC-CPI (ref. 13). A controlled molecular weight LARC-CPI has provided adhesive specimens and composites which exhibited high mechanical properties at temperature as high as 232°C.

In the area of poly(arylene ethers), work in our laboratory has concentrated primarily on the incorporation of heterocyclic units within the polymer backbone (refs. 14-16). In general, heterocyclic units in poly(arylene ethers) provides a higher glass transition temperature, higher tensile strength and modulus. Poly(arylene ethers) containing the following heterocyclic units have been prepared. These polymers

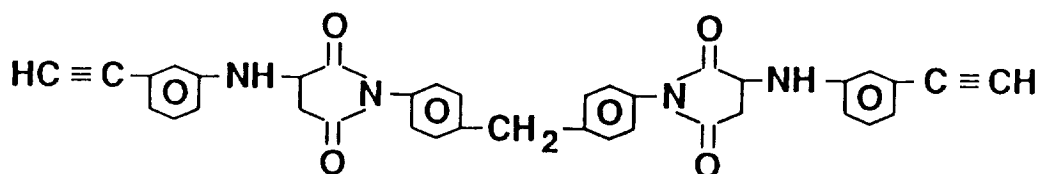
- o Phenyl and substituted phenyl imidazole
- o Phenylquinoxaline
- o Quinoxaline
- o Benzimidazole
- o Benzoxazole
- o 1,3,4-Oxadiazole
- o 1,2,4-Triazole

exhibit a unique combination of properties that classifies them as multipurpose materials. As an example, the properties of a new poly(arylene ether) containing phenylimidazole units as depicted in structure 4 are presented in Table 3.

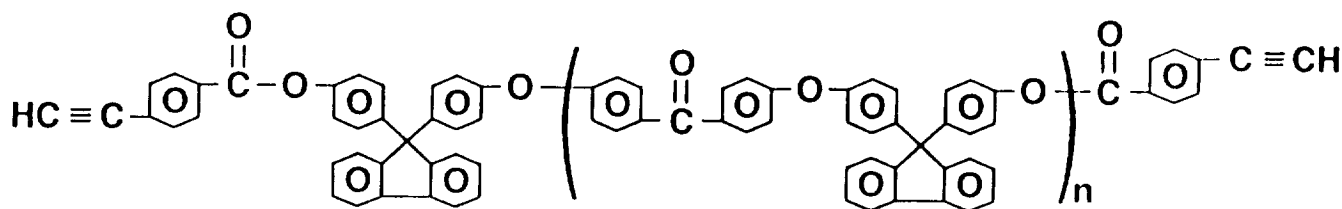


Several poly(arylene ethers) containing heterocyclic units such as the benzoxazole, oxadiazole, quinoxaline and triazole are semi-crystalline. Crystallinity in a polymer generally yields higher modulus and better solvent resistance than an amorphous polymer.

Another approach to the development of high performance/high temperature polymers has involved the use of the ethynyl or acetylenic group. This group has been placed on the ends of molecules, pendent along the polymer chain and within the polymer backbone. Upon heating, it undergoes a complex reaction leading to chain extension, branching and/or crosslinking. In some of our work on ethynyl containing materials, a blend consisting of a low molecular weight ethynyl terminated aspartimide (brittle component) and an ethynyl terminated arylene ether oligomer (tough component) was thermally cured to yield a resin that exhibits attractive neat resin, adhesive and composite properties (refs. 17-18). The chemical composition of the components in the blend are shown in structures 5 and 6. The blend exhibited good compression moldability, allowing for the fabrication of neat resin moldings, adhesive specimens and composites at 250°C under 1.4 MPa (200 psi). The properties of the moldings, adhesive specimens and laminates are given in Table 4.



5



6

### CONCLUDING REMARKS

The high performance/high temperature polymers discussed in this paper are representative of the type of work underway at the NASA Langley Research Center. Further improvement in these materials as well as the development of new polymers will provide technology to help meet NASA future needs in high performance/high temperature applications. In addition, because of the unique combination of properties offered by many of these polymers, they should find use in many other applications.

### REFERENCES

1. J. M. Weatherall and A. T. Ono, paper presented at Am. Chem. Soc. Div. of Polym. Chem. Symposium on "Recent Advances in Polyimides and Other High Performance Polymers," San Diego, CA, Jan. 22-25, 1990.
2. V. L. Bell, B. L. Stump and H. Gager, *J. Polym. Sci. Polym. Chem. Ed.* **14**, 2275 (1976).
3. D. J. Progar and T. L. St. Clair, *Natl. SAMPE Tech. Conf. Series* **7**, 53 (1975).
4. D. J. Progar, V. L. Bell and T. L. St. Clair, U.S. Pat. 4,065,345 (1977) to NASA.
5. A. K. St. Clair and T. L. St. Clair, *Sci. Adv. Mat. Proc. Eng. Series* **26**, 165 (1981).

6. C. L. Hendricks and T. N. Hale in Welding, Bonding and Fastening 1984 Sym. Proceedings, NASA Conf. Publ. 2387 (1985), p. 351.
7. M. Ohta, S. Tamia, T. W. Towell, N. J. Johnston and T. L. St. Clair, Sci. Adv. Matl. Proc. Eng. Series 35, 1030 (1990).
8. D. J. Progar, T. L. St. Clair and J. R. Pratt in Polyimide: Materials, Chemistry and Characterization (ed. C. Feger, M. M. Khojasteh and J. E. McGrath) Elsevier Sci. Publ. B. V., Amsterdam, 1989, p. 151.
9. J. R. Pratt, T. L. St. Clair and D. J. Progar, U.S. Pat. 4,937,317 (1990) to NASA.
10. P. M. Hergenrother, N. T. Wakelyn and S. J. Havens, J. Polym. Sci.: Pt. A: Polym. Chem. **25**, 1093 (1987).
11. P. M. Hergenrother and S. J. Havens, U.S. Pat. 4,820,791 (1987) to NASA.
12. P. M. Hergenrother and S. J. Havens, SAMPE J. **24** (4), 13 (1988).
13. P. M. Hergenrother, M. W. Beltz and S. J. Havens, Sci. Adv. Matl. Proc. Eng. Series 34, 963 (1989).
14. J. W. Connell and P. M. Hergenrother, Polym. Prepr. **29** (1), 172 (1988).
15. J. W. Connell and P. M. Hergenrother, IBID **35**, 432 (1990).
16. P. M. Hergenrother, J. W. Connell and P. Wolf, 33rd Intl. Union of Pure and Applied Chemistry Sym. on Macromolecules, 1990, Session 1.5.2.
17. J. W. Connell, P. M. Hergenrother and S. J. Havens, High Perf. Polym. J. **1** (2), 119 (1989).
18. P. M. Hergenrother, J. W. Connell and S. J. Havens, U.S. Pat. 4,889,912 (1989) to NASA.

Table 1 - Properties of LARC-TPI

Glass transition temperature: ~ 250°C  
 Density: 1.33 g/cc  
 Fracture energy ( $G_{Ic}$ ): ~ 2600 J/m<sup>2</sup> (15 in lb/in<sup>2</sup>)

Unoriented Thin Film Properties at 25°C (dried through 1 hr @ 300°C)<sup>1</sup>

Tensile Strength, MPa (Ksi)	13.6 (19.7)
Tensile Modulus, GPa (Ksi)	3.72 (540)
Elongation, %	4.8

Ti/Ti Adhesive Properties [RT → 343°C under 1.4 MPa (200 Psi)]

Test Condition	Tensile Shear Strength, MPa (Psi)
25°C	42.6 (6000) <sup>2</sup>
232°C	13.8 (2000) <sup>3</sup>
232°C after 1000 hr @ 232°C	15.2 (2200) <sup>3</sup>
232°C after 10000 hr @ 232°C	24.1 (3500) <sup>3</sup>
232°C after 37000 hr @ 232°C	24.1 (3500) <sup>3</sup>

Unidirectional AS-4 Laminate Properties<sup>4</sup>, (cured through 1 hour @ 350°C under ~ 2.1 MPa (300 Psi))

Test Temperature, °C	Flexural St., MPa (Ksi)	Flexural Mod., GPa (Msi)	Short Beam Shear St., MPa (Ksi)
25	1821 (264)	91.0 (13.2)	124 (18.0)
149	1628 (236)	86.9 (12.6)	81 (11.7)
177	1524 (221)	82.8 (12.0)	71 (10.3)

<sup>1</sup>Ref. 2<sup>2</sup>Ref. 3<sup>3</sup>Ref. 6<sup>4</sup>Ref. 7

Table 2 - Properties of LARC-CPI\*

Glass transition temperature: 222°C  
 Crystalline melt temperature: 350°C  
 Melt viscosity at 395°C at angular frequency of 0.1 rad/sec: 10<sup>5</sup> Pa·sec (10<sup>6</sup> poise)  
 Equilibrium moisture pickup: < 1%  
 Dielectric constant at 1 MHz: 3.1  
 Solvent Resistance: Excellent  
 Fracture Energy ( $G_{Ic}$ ): 6650 J/m<sup>2</sup> (38 in lb/in<sup>2</sup>)

Unoriented Thin Film Tensile Properties (Through 1 hr @ 300°C)

Test Condition	Strength, MPa (Ksi)	Modulus, GPa (Ksi)	Elongation, %
25°C	151.7 (22.0)	4.34 (630)	8.3
25°C after 100 hr soak in 30% aq. NaOH	139.9 (20.3)	4.07 (590)	5.0
177°C	104.8 (15.2)	3.72 (540)	21.1
232°C	35.8 (5.2)	1.69 (245)	76.1
232°C after 100 hr @ 316°C in air	57.9 (8.4)	2.35 (341)	9.6

Ti/Ti Adhesive Properties [RT → 400°C under 6.9 MPa (1000 Psi), hold 15 min @ 400°C]

Test Condition	Tensile Shear Strength, MPa (Psi)
25°C	43.1 (6250)
25°C after 1000 hr @ 232°C	49.1 (7120)
177°C	31.1 (4510)
232°C	4.1 (590)
232°C after 1000 hr @ 232°C	18.9 (2740)
232°C after 100 hr @ 316°C in air	25.3 (3670)

\*Ref. 12

Table 3 - Properties of Poly(arylene ether imidazole) 3\*

Glass transition temperature: 248°C  
Fracture Energy ( $G_{IC}$ ): 4000 J/m<sup>2</sup> (22.9 in lb/in<sup>2</sup>)

Unoriented Thin Film Properties

Test Temp., °C	Tensile St., MPa (Ksi)	Tensile Mod., GPa (Ksi)	Elong., %
25	9.8 (14.2)	2.8 (407)	6.0
177	5.7 (8.2)	2.1 (306)	6.0
200	4.6 (6.6)	1.9 (273)	7.5

Ti/Ti Adhesive Properties [RT → 300°C under 1.4 MPa (200 psi)]

Test Temperature, °C	Tensile Shear Strength, MPa (Psi)
25	33.1 (4810)
93	26.2 (3800)
177	25.5 (3700)
200	21.0 (3050)

\*Ref. 15

Table 4 - Properties of Resin from a Blend of Ethynyl Terminated Materials\*

Glass transition temperature: 245°C  
Fracture Energy ( $G_{IC}$ ): 675 J/m<sup>2</sup> (3.8 in lb/in<sup>2</sup>)  
Tensile St.: 93.1 MPa (13.5 Ksi)  
Tensile Mod.: 3.6 GPa (525 Ksi)  
Elong. (break): 2.6%  
Coef. Therm. Expansion: 36.8 ppm/°C

Ti/Ti Adhesive Properties [RT → 250°C under 1.4 MPa (200 psi), hold 0.5 hr]

Test Condition	Tensile Shear St., MPa (Psi)
25°C	22.8 (3300)
25°C after 650 hr @ 200°C, air	17.9 (2600)
150°C	17.2 (2500)
150°C after 650 hr @ 200°C, air	20.5 (2975)
177°C	8.3 (1200)
177°C after 650 hr @ 200°C, air	18.6 (2700)

Unidirectional AS-4 Laminate Properties (cured 1 hr @ 250°C under 1.4 MPa)

Test Temp., °C	Flexural St., MPa (Ksi) [Mod., GPa (Msi)]	Tensile St., MPa (Ksi) [Mod., GPa (Msi)]
25°C	1517 (220) [104 (15.1)]	1917 (278) [121 (17.6)]
150	1462 (212) [100 (14.5)]	1737 (252) [139 (20.2)]
177	1359 (197) [97.2 (14.1)]	—

\*Ref. 17

## INDUSTRIAL APPLICATIONS OF GRAPHITE FLUORIDE FIBERS

Ching-eh Hung  
NASA Lewis Research Center  
Cleveland, OH 44135

and

Donald Kucera  
Cleveland State University  
Cleveland, OH 44115

## SUMMARY

Based on fluorination technology developed during 1934-1959, and the fiber technology developed during the 1970's, a new process was developed to produce graphite fluoride fibers. In the process, pitch based graphitized carbon fibers are at first intercalated and deintercalated several times by bromine and iodine, followed by several cycles of nitrogen heating and fluorination at 350-370°C. Electrical, mechanical, and thermal properties of this fiber depend on the fluorination process and the fluorine content of the graphite fluoride product. However, these properties are between those of graphite and those of PTFE (Teflon). Therefore, it is considered to be a semiplastic. The unique physical properties suggest that this new material may have many new and unexplored applications. For example, it can be a thermally conductive electrical insulator. Its coefficient of thermal expansion (CTE) can be adjusted to match that of silicon, and, therefore, it can be a heat sinking printed circuit board which is CTE compatible with silicon. Using these fibers in printed circuit boards may provide improved electrical performance and reliability of the electronics on the board over existing designs. Also, since it releases fluorine at 300°C or higher, it can be used as a material to store fluorine and to conduct fluorination. This application may simplify the fluorination process and reduce the risk of handling fluorine.

## INTRODUCTION

Graphite fluoride is a product of elevated temperature fluorination of graphite. The graphite fluoride powder has been studied extensively for several decades since it was first reported in 1934 (Ref. 1, 2). The original objective of the present effort was to use this old knowledge of graphite fluorination, together with the new pitch-based graphitized carbon fibers (developed in the 1970's), to produce graphite fluoride fibers. These fibers could then be processed with the currently available fiber technology to make thermally conductive and electrically insulative products, which are useful in electrical or electronic industries for heat sinking purposes.

During the process of conducting this research, in order to reach the objective described above, the old knowledge of graphite fluorination needed to be augmented, and further development of the newly developed carbon fibers is needed. Also, based on the electrical, mechanical, and thermal properties of the graphite fluoride fibers thus obtained, many applications in addition to heat sinking are believe possible. In this report, the fluorination process and its effects on the physical properties of the graphite fluoride fiber products will be summarized, and the applications of the graphite fluoride fibers thus produced will be described.

## GRAPHITE FLUORINATION PROCESS

Results of the study on graphite fluorination during 1934-1959 indicated that after contacting fluorine gas at 400-460°C temperature range, the planar molecular structure of graphite crystal (Figure 1a) became buckled, and carbon atoms were bonded to fluorine atoms covalently (Figure 1b) (Ref. 1, 2).



Figure 1b shows the molecular structure of a perfect graphite fluoride crystal whose fluorine to carbon ratio is 1.0. Such structure is seldom produced either in industry or in laboratory. This is because that, with direct contact of graphite crystals, the fluorine gas will either react only with the carbon atoms near the exterior region of the crystal at low temperature, or break the crystal structure in order to reach the interior carbon atoms at high temperature. Consequently, the graphite fluoride thus produced has less fluorine than carbon, and is structurally damaged.

In order to alleviate these problems, the graphite materials to be reacted with fluorine at high temperature were pre-tested with chemicals, called intercalants, which interleave between the carbon layers, to form an intercalated compound. The molecular structure of a typical intercalated compound is described in Figure 1c, in which bromine, the intercalant, is inserted into the space between the graphite planes. It is believed that the intercalants help to "open" the structure of the graphite material for fluorine to reach the carbon atoms near the center of the graphite crystal. This concept has been experimentally demonstrated (Ref. 3, 4).

The above concept suggests that the quality of the final graphite fluoride products depends not only on the graphite fluorination process, but also on the graphite pre-treatment process, or intercalation. The intercalation needs to be as complete as possible in order to eliminate pockets or pristine graphite, which would be the sources of structural damage, or defects, in the graphite fluoride fiber products. Also, the choice of intercalants needs to be such that it causes the least structural damage during intercalation. From the knowledge of graphite intercalation accumulated during the past 7 years of research at NASA Lewis Research Center, it was determined that bromine and iodine are the best intercalants for this purpose (Ref. 5, 6). Details on graphite intercalation with bromine and iodine are described in references 6 and 7.

The reaction of fluorine at 1 atm pressure and graphitized carbon fibers has been studied detail (Ref. 4, 9). The fluorine used in this work was pure commercially available fluorine containing 2-3%  $N_2$  and HF. The effects of this impurity and the fluorine pressure on the fluorination process are not known. However, 1 atm HF was reported to react with graphite at 250° C (Ref. 2). It may act as a catalyst in the fluorination reaction. The fibers used in this work were mostly Amoco P-100 graphitized carbon fibers which contain 18% bromine by weight after bromine treatment. The P-100 fiber has interplanar spacing of 3.37 Å before bromine intercalation and 3.39 Å after bromine intercalation. Less graphitized carbon fibers were also used in a few experimental runs. This study concluded that the optimum reaction temperature is in the 350-370° C range for brominated P-100 fibers, and lower for less graphitized fibers. Also, the time needed for complete fluorination can be reduced if 2 or more cycles of nitrogen heating and fluorination at this temperature are conducted instead of a continuous fluorination process. It is believed that the nitrogen heating process helps to clear the pathway for fluorine to reach the carbon atoms near the center of the graphite crystals.

## PROPERTIES OF GRAPHITE FLUORIDE FIBERS

The work done during 1934-1959 concluded that graphite fluoride was an electrical insulator having fluorine to carbon ratio of 1.0. However, the "complete fluorination" described in the last paragraph produced an electrically insulative fiber having fluorine to carbon atom ratio of 0.68 to 0.72. Further fluorination results in more structural defects but little increase in fluorine content in the products. The fact that the graphite fibers described here are electrical insulators but have much less fluorine than carbon suggests the formation of new carbon-carbon covalent bonds between the carbon atoms from the neighboring graphite layer (Ref. 8). However, it may also result from isolate double bond regions which could not be reached by fluorine during the fluorination reaction.

Some properties of the  $CF_{0.68}$  graphite fluoride fibers were measured and are described in Table 1. It is noted that, with further optimization, the structural damage to the fibers during the fabrication process could be further reduced. Therefore, the thermal conductivity and the tensile strength of the fibers could increase.

The properties of pristine graphite fibers, graphite fluoride fibers, and fiber glass are compared in Table

1. It shows that the fiber properties change significantly as the fluoride content in the fibers changes from 0 to 0.68. Using the process described, partially fluorinated graphite fibers having fluorine content between 0 and 0.68 were also made. Further work concluded that partially fluorinated graphite fibers have physical properties between those of the pristine graphite fibers and those of the  $CF_{0.68}$  fibers described in Table 1. The effect of the fluorine content on the fiber resistivity and the Young modulus are described in Reference 9. Quantitative study of the effect of fluorine content on the fiber CTE is time consuming and is not completed at this time. However, a clear trend of increasing CTE value with increasing fluorine content is observed. Also, after repeated experimental tests, the CTE value for the  $CF_{0.68}$  graphite fluoride fibers is determined to be in the 2-7 ppm/ $^{\circ}$  C range. For comparison, the pristine graphitized carbon fiber has a CTE value of -1 ppm/ $^{\circ}$  C.

The physical properties of  $CF_{1.0}$  described by previous scientists (Ref. 1, 2) were somewhat different from those of the  $CF_{0.68}$  fibers described in Table 1. This again demonstrates that the physical properties of the graphite fluoride are strongly affected by the fluorine content in the graphite materials.

On the other hand, the physical properties of  $CF_{1.0}$  are similar to those of PTFE (Teflon), whose chemical composition is approximately  $CF_2$ .

Based on the experimental results described above, partially fluorinated graphite fluoride described in the report can be considered as a semiplastic.

Figure 2 shows the % weight loss after a  $CF_{0.68}$  fiber sample was heated in air at different temperatures for 17, 24, and 192 hours. It shows that the weight of this graphite fluoride fiber sample changes very little under 200 $^{\circ}$  C air, but decreases slowly and continuously if the air temperature is 300 $^{\circ}$  C or higher. Data from ESCA and SEM's EDS suggest that the small amount of weight loss at 200 $^{\circ}$  C is not the result of loss of fluorine, but rather is due to the loss of a small amount of oxygen impurity in the fiber. On the other hand, the weight loss at 300 $^{\circ}$  C or higher is largely due to the loss of fluorine in the fibers (Ref. 10).

The electrical resistivity of graphite fluoride fibers, however, decreases from around 11 $^{11}$   $\Omega$ -cm to around 10 $^6$   $\Omega$ -cm. However, the graphite fluoride fibers do not react with anhydride-type epoxy hardener. The electrical resistivity of a composite made from such epoxy is found to be stable at room temperature and pressure, and is in the same range as that of the graphite fluoride fibers which made the composite.

## APPLICATIONS OF GRAPHITE FLUORIDE FIBERS

The application of graphite fluoride fibers in space industry are summarized in Reference 11. Those applications are based on the unique properties of the graphite fluoride fibers described above. These properties, which can best be described as semiplastic, can certainly be used in other industries as well. As with any new fiber material, many new applications of the graphite fluoride fibers and its composites remain to be explored. However, some examples of the nonaerospace applications are described as follows.

### Printed Circuit Board

Traditionally, the material used for printed circuit boards is fiberglass epoxy composites. Typically, it has CTE value of 7-13 ppm/ $^{\circ}$  C, and thermal conductivity of 0.3-0.4 W/M- $^{\circ}$ K.

The disadvantage of this CTE value is that it is somewhat different from the CTE value of silicon (3-7 ppm/ $^{\circ}$  C). The CTE mismatch generates stress on the pins of the silicon chip. As a result, if the device is operated over a wide temperature range, the joints between the pins and the board may crack, and the device becomes unreliable. This is especially true in the more advanced designs, where the silicon chips are large.

The disadvantage of the above described thermal conductivity value for fiberglass composites is that it is very low. Therefore, heat generated from conducting electricity through the electronics on the circuit boards

cause elevated device temperatures. Such high temperatures affects not only the electrical performance, but also the reliability because it enhances chemical corrosion, interfacial diffusion, and again, CTE mismatch (Ref. 12). This problem is especially true for the more advanced devices, where a large number of electronic components are packed in a small area.

The above disadvantages may be partially alleviated by using graphite fluoride fibers to replace fiberglass. For unidirectional graphite fluoride fiber-epoxy composites, the CTE and thermal conductivity values are estimated to be in the 3-6 ppm/ $^{\circ}$ C and 3-6 W/M- $^{\circ}$ C ranges, respectively, along the fiber directions. These values are better than those of fiberglass composites. In addition, further optimization of the fluorination process should result in even better CTE and thermal conductivity values.

Adhesion between metal and graphite fluoride is a concern for the use of graphite fluoride composites as a printed circuit board material. This property has not been studied in detail. However, letting a strand of graphite fluoride fibers, CF<sub>0.65</sub>, be sandwiches between, and in direct contact with, two pieces of commercially available titanium foil at 200 $^{\circ}$ C for 2 hours, some fibers were found to adhere to the foils. This suggests that adhesion between graphite fluoride and titanium is possible. Using aluminum as the foil material and repeating the same experiment, the same phenomenon was observed, but to a lesser extent. The adhesion behavior observed here seems to be similar to that between metal and fluorocarbon as observed by Kid, et al (Ref. 13). However, more work is needed to understand and characterize this phenomenon.

Another factor important to the design of printed circuit boards made from graphite fluoride fibers is that the dielectric constant of the board needs to be low in order to allow signals to travel fast. The dielectric constant of graphite fluoride fibers, however, is not known at this time. However, it is believed that the impurities in the graphite fluoride fibers, i.e., oxygen, bromine, and iodine, need to be removed in order to have a product which contains no "conductive pockets" and, therefore, which has a minimum value of dielectric constant.

#### Fluorine Storage and Fluorination Agent

As described in Figure 2, partially fluorinated graphite fluoride losses fluorine if heated at 300 $^{\circ}$ C or higher. This prevents such graphite fluoride from being used at high temperatures. However, this also suggests that graphite can be used as a host material to store fluorine, and the partially fluorinated graphite fluoride can be used as a fluorination agent. Large amounts of highly reactive and toxic fluorine gas can be stored in the form of a solid, e.g., spools of graphite fluoride fibers which typically contain 50% fluorine by weight is easy to handle at room environment. Fluorine gas can then be released for use from this graphite fluoride by heating it at 300 $^{\circ}$ C or higher. The fluorine release rate can be controlled by adjusting the graphite fluoride heating temperature. An example of the possible use of this process in laboratory is to study the effects of fluorination on high temperature superconductors (Ref. 14). However, application of this process in industry, e.g., etching of amorphous silicon and silicon nitride thin films (Ref. 15), remains to be explored.

#### Other Applications

Some of the other applications of this new material that the authors believe possible are listed as follows:

- A lubricant (Ref. 16)
- The cathode material in lithium battery (Ref. 17)
- Insulators for windings of alternators
- Filler material which bonds to metal matrices to form a composite with strong traverse tensile strength

#### CONCLUSION

Graphite fluoride fibers with a fluorine to carbon atom ratio of 0.68 or lower were fabricated at NASA

Lewis Research Center. Physical properties of this material can be controlled to some extent by adjusting its fluorine content. For  $CF_{0.66}$ , the material behaves like a semiplastic. As with any new material, applications of the graphite fluoride fibers remain to be explored. However, it appears promising to use this material to fabricate heat sinking printed circuit boards which are CTE compatible with silicon. It also seems possible to use the graphite fiber as a host material to store fluorine gas, and to use the graphite fluoride fiber a safe fluorination agent.

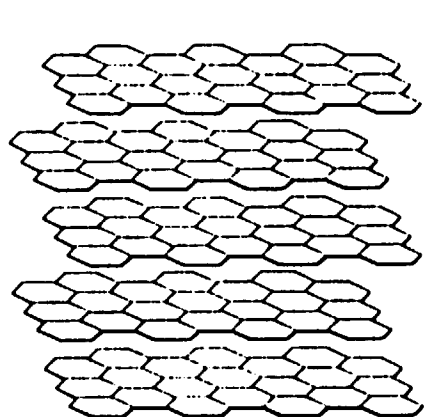
## REFERENCES

1. Ruff, O., and Bretschneider, O.: Neues Aus Der Chemie Des Fluors. Z. Anorg. U. Allgem. Chem., Vol. 217, p. 1, 1934.
2. Rudoff, W.: Graphite Intercalation Compounds, in Advances in Inorganic Chemistry and Radiochemistry, Vol. 1, p. 244, 1959.
3. Watanabe, N., Nakajima, T., and Kawaguchi, M.: Process for Producing a Graphite Fluoride Comprising Mainly Polydicarbon Monofluoride Represented by the Formula  $(C_2F)_n$ ; Reaction of Graphite Material with Fluorine in the Presence of a Fluoride. U.S. Patent 4,423,261(Dec. 1983).
4. Hung, C., Long, M., and Stahl, M.: Synthesis, Electrical and Thermal Conductivities and Potential Applications of Graphite Fluoride Fibers. SAMPE Q., Vol. 19, p. 12, 1988.
5. Hung, C., and Miller, J.: Thermal Conductivity of Pristine and Brominated Graphite Fibers. Carbon, Vol. 25, No. 5, p. 679, 1987.
6. Hung, C., and Kucera, D.: Intercalated Graphitized Carbon Fibers with Iodine as the Major Intercalant. To be presented at The 20th Biennial Conference on Carbon, The University of California at Santa Barbara, June 23-28, 1991.
7. Jaworske, D., Gaier, J., Hung, C., and Banks, B.: Properties and Potential Applications of Brominated P-100 Carbon Fibers. SAMPE Q., Vol. 18, p. 9, 1986.
8. Touhara, H., Kadono, K., Fujii, Y., and Watanabe, N.: On the Structure of Graphite Fluoride. Z. Anorg. Allgem. Chem. Vol. 544, p. 7, 1987.
9. Hung, C., and Long, M.: Fluorinated Graphite Fibers as a New Engineering Material: Promises and Challenges. NASA TM-102511, 1989.
10. Hung, C.: Heating Effects on Graphite Fluoride Fibers. To be presented at The 20th Biennial Conference on Carbon, The University of California at Santa Barbara, June 23-28, 1991.
11. Hung, C., Long, M., and Dever, T.: Graphite Fluoride Fibers and Their Applications in Space Industry. NASA TM-103265, 1990.
12. IEEE Videoconferences Seminars via Satellite. High Performance Integrate Circuit Packaging. September 22, 1987.
13. Kid, Y., Chang, C., and Schrott, A.: Adhesion of Metals to Spin-Coated Fluorocarbon Polymer Films. J. Appl. Phys. Vol. 67, p. 251, 1990.
14. Perrin, C., Pena, O., Dinia, A., Sergent, M., Fonteneau, G., Christensen, P., Lucas, J., Burlet, P., and Rossat-Mignod, J.: Actual Fluorination of  $YBa_2Cu_3O_x$  Superconductors: Enhancement of Superconducting Properties and Neutron Diffraction Studies. Physica C, Vol. 162-164, p. 889, 1989.

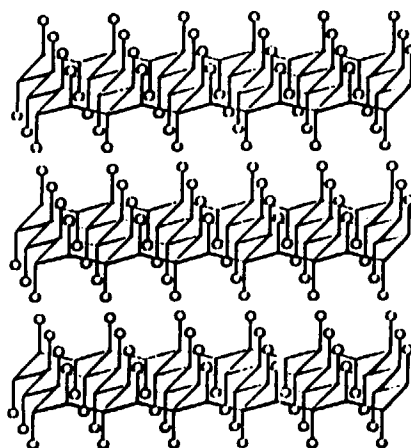
15. Kuo, Y.: Reactive Ion Etching of PECVD Amorphous Silicon and Silicon Nitride Thin Films with Fluorocarbon Gases. J. Electrochem. Soc., Vol. 137, p. 1235, 1990.
16. Fusaro, R.: Graphite Fluoride Lubrication: The Effect of Fluorine Content, Atmosphere, and Burnishing Technique. ASLE Transactions, Vol. 20, p. 15, 1976.
17. Watanabe, N., and Nakajima, T.: Graphite Fluoride, in Preparation Properties and Industrial Applications of Organofluorine Compound, ed. Banks, R., p. 297, John Wiley and Son. 1982.

Table 1 Physical properties of graphitized carbon fibers (P-100), graphite fluoride fibers ( $\text{CF}_{0.68}$ , made from P-100 fiber), and fiber glass (S type)

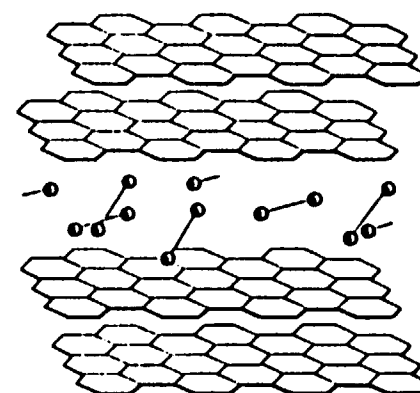
	Graphite fiber	Graphite fluoride	Fiber glass
Electrical resistivity ( $\Omega\text{-cm}$ )	$2.5 \times 10^{-4}$	$10^{11}$	$10^{14}$
Thermal conductivity ( $\text{W/M-}^\circ\text{K}$ )	300	11	1.1
Young modulus (Msi)	105	25	12
Longitudinal tensile strength (Ksi)	300	40	500
Coefficient of thermal expansion ( $\text{ppm/}^\circ\text{C}$ )	-1	2-7	3
Density ( $\text{g/cm}^3$ )	2.18	2.5	2.5



A) GRAPHITE



B) GRAPHITE FLUORIDE



C) BROMINE INTERCALATED GRAPHITE

FIGURE 1 - MOLECULAR STRUCTURES

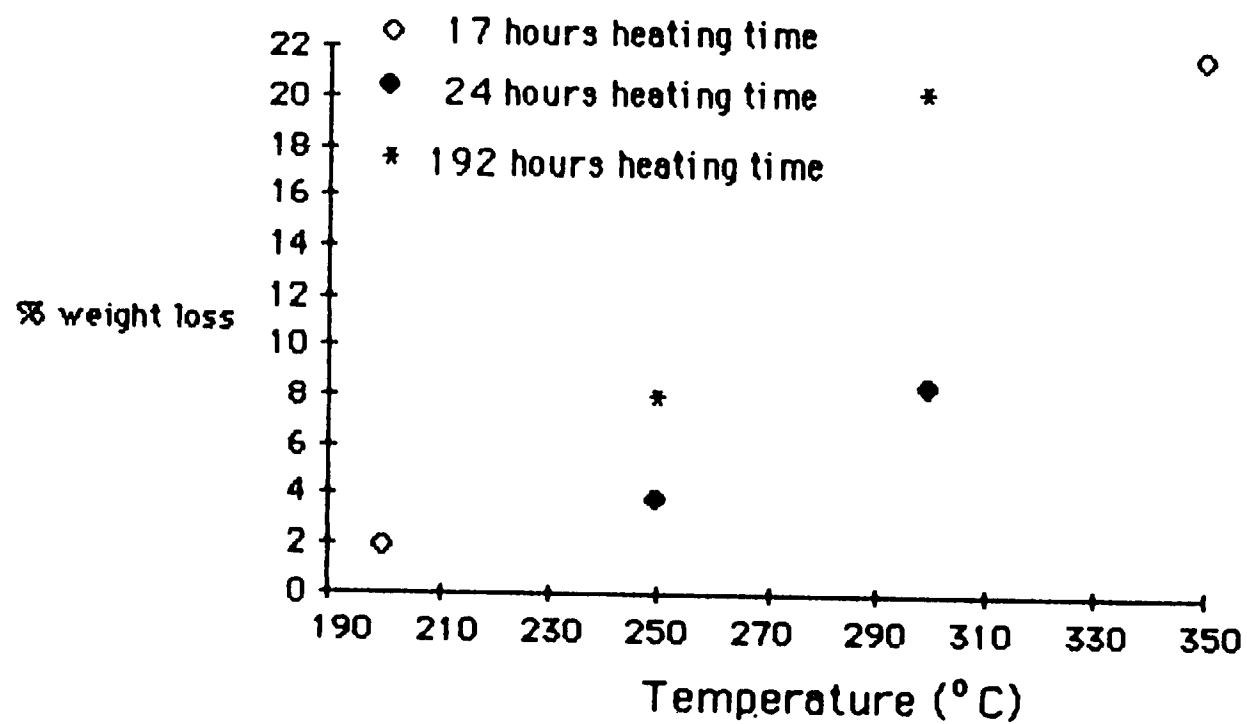


Figure 2 Weight loss after heating  $CF_{0.65}$  in air



## UNIQUE APPLICATIONS OF FLUOROEPoxy MATERIALS

Sheng Yen Lee  
NASA Goddard Space Flight Center

## INTRODUCTION

Teflon has become essential in our daily life both at home and in industry. The trade name Teflon symbolizes a family of fluorocarbon thermoplastics that are known to have high thermal stability, good low temperature properties, low surface energy, low moisture absorption, excellent electrical properties, and particularly chemical inertness. Just because of its inertness and its high thermal stability, Teflon processing is very difficult. There is no good method to cure or crosslink Teflon yet.

It has been a challenge to develop a fluorinated thermoset material with properties similar to Teflon yet can be processed like regular thermosets. Let's imagine how nice it would be if our cars or boats could be spray-painted with Teflon paint. Our war ships would last longer at much lower maintenance cost and move faster without the nuisance of barnacle problem in the sea if they could be painted with Teflon-like epoxy paint. Indeed, Griffith and his coworkers in the U.S. Navy have worked hard toward this goal and have developed a basic fluorodiol structure from which a series of fluoroepoxies and fluoropolyurethanes can be prepared (1, 2). Although there is still a long way before we can see U.S. war ships painted with a fluoroepoxy paint, now fluoroepoxy resin may be commercially available in pilot-plant quantities, and its attractive potential can be demonstrated by a few unique applications presented in this paper.

Fluoroepoxy and curing agents. We at NASA are interested in this new material. We believe that fluoroepoxy resin could be formulated to make a good moisture vapor barrier coating. The first problem in formulating a fluoroepoxy material is to find a good compatible curing agent. Few fluorinated epoxy curing agents are commercially available. Figure 1 shows the fluoroepoxy resin I developed by the Navy and a few curing agents available that have been tried.

Fluoroepoxy I is a diepoxide functionally identical to the conventional epoxy resin. Its fluorine content may vary with the perfluoroalkyl group,  $R_f$ . For this work, the highly fluorinated epoxy C8F was generally used unless indicated otherwise. Fluorinated anhydride II was synthesized by the Naval Research Laboratory. Like other anhydrides, it needs elevated temperature curing. Siloxane diamine III, supplied by Silar Laboratories, Scotia, New York, could be made compatible with I-C6F at 50° C, but the system is handicapped by the high sensitivity of the amine to carbon dioxide in air as well as by the concern that the siloxane structure may compromise the low moisture permeation requirement. Fluorinated diamines IV and V are laboratory products. Unfortunately they are either sluggish or inactive in curing epoxy I. Polyamine VI is an adduct amine that can be conveniently prepared by reacting epoxy I-C6F with an excess of ethylenediamine. The unreacted diamine can be removed easily due to its relatively low boiling point. Curing with the adduct amine is equal to curing with ethylenediamine except that the ethylenediamine is pre-tied to epoxy I to minimize the compatibility problem. The liquid mix of epoxy I and adduct amine VI appears cloudy but compatible, and the cured products are transparent. The adduct curing agent may be less than perfect for many applications where more flexibility of the cured material is desired. Nevertheless, it gives room temperature gelation and the cured products can pass the stringent outgassing requirement that NASA emphasizes for space uses. Therefore, this paper presents applications basically using the adduct amine VI as the curing agent.

An excellent moisture vapor barrier coating. Fluoroepoxy I-C8F and adduct amine VI were formulated in stoichiometric amounts. For spray-coating, solutions of ca. 50% by wt. were made with a solvent mix of 1,1,1-trichloroethane, THF, methylene chloride, and methanol in a ratio of 20/20/30/30 by volume. Spraying was done with compressed helium as a carrier gas to avoid gas trapping as well as the appearance of orange peel which was observed when nitrogen was used. After

room temperature gelation, curing was completed by heating at 80° C for 2 hours.

A low moisture vapor transmission rate is the key to a good moisture vapor barrier material. For the determination of the transmission rate, fluoroepoxy films were prepared by spray coating. Solvent trapping was a problem, a dilemma expected with a good vapor barrier material. To reduce trapping, a quite volatile solvent system was developed as given above. Film thickness was built with multiple light sprayings. Time was allowed for solvent evaporation between sprayings, and the curing temperature was programmed to facilitate the solvent release. With these precautions, it was possible to keep the 125° C vacuum outgassing level of the cured film at about 1.5% total weight loss. The level could be brought down to 1% or less by baking the film at 100° C in order to meet the NASA requirement.

The moisture transmission rates of the fluoroepoxy film and the films of a number of reference materials were determined according to an ASTM method. The data collected are shown in Table 1. Generally, the rates are inversely proportional to the fluorine content of the materials. The transmission rate of the fluoroepoxy is higher than that of the highly fluorinated Teflon PFA and Teflon FEP, but is only about one twelfth of that of the four polyurethane coatings/potting materials popular in the market. Uralane and Conathane EN11 are polybutadiene-diol based polyurethane and are noted for their excellent moisture resistance.

Closely related to moisture transmission is moisture absorption. Three materials were compared in the moisture absorption experiment conducted at room temperature and 97% relative humidity (RH). The results shown by Figure 2 indicate the overwhelming superiority of the fluoroepoxy and polyurethane EN11 over epoxy Epon815/V140 (Epon815 and Versamid 140 in 50/50 ratio by wt.). Plotting on a different scale in Figure 3 illustrates that the moisture absorption of the fluoroepoxy had essentially leveled at 0.26% in 27 weeks, while that of EN11 was approximately two times higher and the difference increases continuously with time.

As adhesives to bond Teflon without any surface treatment. Teflon is non-polar, non-wettable, and consequently non-bondable with adhesives if its surface is not pre-treated. Since Teflon is noted for its chemical inertness, some special methods have to be used for the surface treatment. Up to now, the primary method is dependent on a special chemical etching agent which is a solution containing sodium-naphthalene complex. However, the reagent is corrosive and may be hazardous in use. In many cases, it is undesirable or simply impossible to conduct such an etching operation on an object to be bonded. Moreover, the etching action has to be drastic so that the Teflon surface can be converted from being non-polar to polar.

After the conversion, the surface layer is so much different from the bulk of the polymer that the molecular bond of the etched layer to the bulk of the plastic is seriously weakened. In fact, the layer becomes a weak interface subject to moisture penetration and layer separation even when it is bonded with a superb adhesive. The sensitivity of the etched layer to moisture makes it behave like an effective humidity sensor (4). Its sensitivity to air, oxygen, and ultraviolet light makes it necessary that either the etched Teflon be bonded without delay or be sealed from air and light in order to avoid rapid degradation. In view of all these shortcomings, people in the field have long sought after an adhesive that can bond Teflon or other fluoroplastics without any special surface treatment.

"Like dissolves like" is the old simple rule. The fluoroepoxy, like other epoxy resins, may be used as an adhesive, and especially it may be good to bond Teflon which is also a fluorinated plastic. To test its effectiveness, aluminum and Teflon PTFE rod adherends were made according to an ASTM method. One set was bonded with the fluoroepoxy adhesive formulated stoichiometrically with fluoroepoxy I-C8F and adduct amine VI, while the other set was bonded with a popular epoxy adhesive Epon828/V140 (50/50 by wt.) for comparison. The adhesive tensile strengths shown in Table 2 demonstrate that to bond aluminum, the strength of the fluoroepoxy is only 60% of that of Epon828/V140, yet to bond Teflon, its strength is 2 times that of the latter. The sodium naphthalene

etching of Teflon improves the bond strength of the Epon828/V140, but only to a level 40% higher in average than that achieved by the fluoroepoxy without etching. The strength gained by the etching varied greatly with the degree of etching, and the etching led to a bond failure by separation of the etched surface layer. In contrast, the fluoroepoxy adhesive offers a new convenient way to bond Teflon without etching to achieve a considerable bond strength free from risk.

To verify the effect of the fluorine content, fluoroepoxy adhesives were prepared with fluorine content varied from 56.4 to 45.9% by using fluoroepoxy I-C8F, -C6F, and -C0F, with adduct amine VI as the curing agent in stoichiometric amounts. The adhesive tensile strengths obtained in bonding Teflon PTFE without etching are given in Table 3. The trend shown is that the strengths decreases only slightly with the decrease of the F-content of the adhesives until the content drops to a level of around 46% where the bonding strength falls sharply to a value as low as that of the common adhesive containing no fluorine. It is important to note that here the 46% marginal level is for bonding Teflon PTFE which is a fully fluorinated plastic with the highest F-content of 76%. We have evidences to indicate that the novelty of the fluoroepoxy adhesives is limited to the application to the highly fluorinated plastics. For bonding Tefzel, a random copolymer of ethylene and tetrafluoroethylene made by DuPont with a F-content of 54%, there is little advantage to choose fluoroepoxy adhesives over a conventional adhesive (5, 6).

A new method to make thermosetting fluoropolymer foam. Teflon has been widely used for electrical insulation and as structural components due to its excellent electrical and other properties. For many applications, Teflon and other thermoplastic fluoropolymer foams are particularly desirable because they render reductions in weight, dielectric constant, and dissipation factor. There are many patented processes to make Teflon foam. However, all of them require a minimum temperature of 240° C and a pressure of up to 1500 psi or much higher. These methods employ volatile or chemical blowing agents and nucleating agents which may leave undesirable residues in the foam products. The main reason for the required extreme conditions is that the available thermoplastic fluoropolymers are processable only at such high temperatures and the high temperature dictates the use of high pressure to control foaming. Consequently, these requirements can constitute a prohibitive barrier to many applications.

With liquid fluoroepoxy resins, regular epoxy foaming processes should be practical. Particularly a new foaming method has been developed on the basis of a novel property of the fluoroepoxy observed. The highly fluorinated epoxy compound made of I-C8F and adduct amine VI is viscous but has minimal tendency to trap the air introduced in mixing the two components. In other words, the usual vacuum deaeration step following the mixing step is easy and fast. Just as it is easy to get the air out under vacuum, gas can also be readily dispersed in the fluoroepoxy compound under a moderate pressure. Perhaps this may be attributed to the low molecular forces between fluorinated molecules. It is this unusual property that offers a new foaming process for fluoroepoxy as well as other thermosetting fluoropolymeric materials.

The process is simple. A chosen unreactive gas is dispersed at or near room temperature in the fluoroepoxy compound before gelation by exposing the compound to the gas under a moderate pressure. The compound gas will be trapped in the compound upon gelation. The gelled clear compound will then expand to form a foamy or cellular structure when it is heat-cured to its completion (3, 7, 8). The process is contrary to the common pressure potting method, which removes trapped air and eliminates voids in the common potting compound by pressing the compound in a gas chamber under a moderate pressure for a period of time or until gelation.

The depth of the gas dispersion in the fluoroepoxy compound depends on the type of the gas and the magnitude of the gas pressure. The cellular structure may be controlled by the degree of gelation or cure and the temperature programming in the heat-cure step. Generally, a pressure of 3 atmospheres or 45 psi suffices to make a foam sheet 3 mm thick. The foam cell is essentially closed. The fact that one has the freedom to choose any unreactive gas for foaming is practically a freedom to store a

chosen gas in the cellular material. Air and nitrogen may be used for ordinary purposes. Oxygen, argon, krypton, sulfur hexafluoride, and many others may be the choice when some special properties from the gas may be desired. The stored gas will remain in the cell for a long time because of the low permeation rate.

As a new antifoaming agent for epoxy material manufacturing and processing. The control or elimination of the foam that arises in many industrial processes can be a critical factor in their operation and their cost reduction. Antifoaming agents are therefore needed to inhibit and reduce foam formation or to rapidly knock down foam (to defoam) when it occurs. The annual world market for antifoaming agents probably exceeds a quarter of a million metric tons.

Antifoaming agents are widely used in the production of monomers and polymers, in the textile industry, in the paint and latex industry and other polymer coating processes, and in the manufacture and application of adhesives. In our investigation of copolymerizing the fluoroepoxy with Epon828, it was found that the use of I-C8F in a very small amount could effectively break the foam bubbles and cut tremendously the vacuum deaeration time in the preparation of the epoxy compound. Air entrainment almost always occurs when the components of an epoxy compound are mixed either manually or with a mechanical stirrer. The mixing is normally followed with vacuum deaeration to remove the entrained air from the viscous liquid compound. The deaeration process is time-consuming and involves fighting patiently with foaming-up. Any residual air in the compound may be the cause of a weakened bond when used as an adhesive or a composite matrix, or may be an electrical insulation failure when used for embedding or impregnation.

The antifoaming effect of fluoroepoxy I-C8F was tested with both commercial epoxy materials and our laboratory formulated epoxy compounds as shown in Table 4. Evidently, I-C8F shows excellent antifoaming effect when applied at a level of as low as 0.1 to 0.5% by weight in the epoxy compounds .

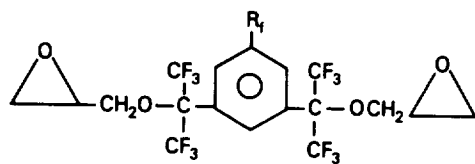
Many antifoaming agents are commercially available, but those which have little side effects are limited for epoxy compounds. The novelty of the fluoroepoxy is that it is reactive and can copolymerize with a conventional epoxy resin at the same time in the same way in forming an inter-crosslinking polymer network. Consequently, the antifoaming agent is not a free additive left in the cured material. Our experimental data indicate that as an antifoaming agent used only in small amount, I-C8F does not have any adverse effect on outgassing property and adhesive strength of either commercial adhesives tested or epoxy compounds formulated in our laboratory. Table 5 shows the adhesive tensile strength data with and without I-C8F added.

It is reasonable to expect that fluoroepoxy resins can be used as an effective antifoaming agent in the early stage of the epoxy manufacturing. It is known that foaming is usually a problem when excess reactants or solvents are stripped. Also, the antifoaming application may not be limited to epoxy processing either. A similar advantage will exist when the fluoroepoxy is applied to polyurethane and other thermoset polymers' processing where the functional groups of the pre-polymers are reactive to the epoxide. Their reaction will also tie up the fluoroepoxy to the crosslinked polymer network.

Conclusion. The four unique applications presented above are just examples to demonstrate the attractive potential of the fluoroepoxy. There may be many other novel properties waiting to be explored. Individually, epoxy resins and fluoropolymers are two known classes of important, wonderful materials. Their "married" product, fluoroepoxy material, is bound to possess novel properties. At this time, fluoroepoxy resin is still an exotic material in terms of cost. But when more novel properties and applications are identified, increased interest in the material may create a market demand that can make the fluoroepoxy cost effective. High tech competition is intensifying as we march toward the twenty-first century. Today's exotic material should be seriously considered in order to meet the future challenge.

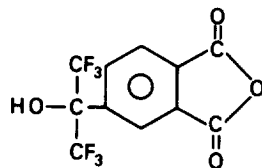
## REFERENCES

1. Griffith, J. R.: Epoxy Resin Containing Fluorine. Chem. Tech. Vol.12, No.5, 1982, PP.290-3.
2. Field, D. E. and Griffith, J. R., U.S. Patent 4 132 681, 1979.
3. Lee, Sheng Yen and Griffith, J. R.: Fluoroepoxy Resin for Moisture Vapor Barrier Coating and Other Applications. I&EC Product Research and Development Vol.25, No.4, 1986, PP.572-7.
4. Drake, S. B., Bhoraskar, S. V., Patil, P. A., and Narasimhan, N. S.: Chemically modified Teflon as an Effective Humidity Detector. Polymer, Vol.27, No.6, 1986, PP.910-2.
5. Lee, Sheng Yen: The Use of Fluoroepoxy Compounds as Adhesives to Bond Fluoroplastics without Any Surface Treatment. SAMPE Quarterly, Vol.19, No.3, April 1988, PP.44-8.
6. Lee, Sheng Y., U.S. Patent 4 902 574, 1990
7. Lee, Sheng Y., U.S. Patent 4 731 211, 1988
8. Lee, Sheng Y., U.S. Patent 4 843 123, 1989

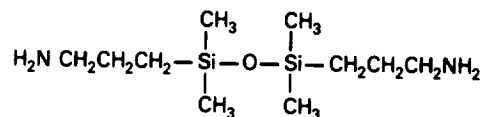


I.

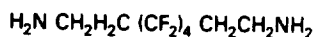
$R_1 = \text{H}$       C0F 43.7% F  
 $= \text{C}_6\text{F}_{13}$     C6F 56.5% F  
 $= \text{C}_8\text{F}_{17}$     C8F 58.6% F



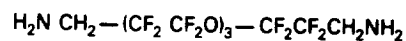
II.



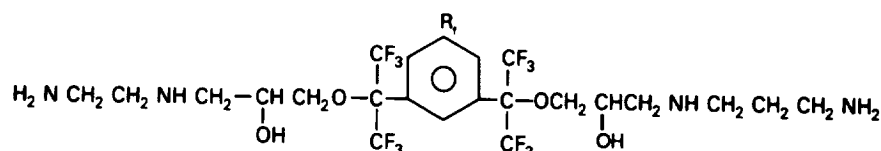
III.



IV.



V.



VI. Adduct Amine

Figure 1. Fluoroepoxy Resins and Curing Agents

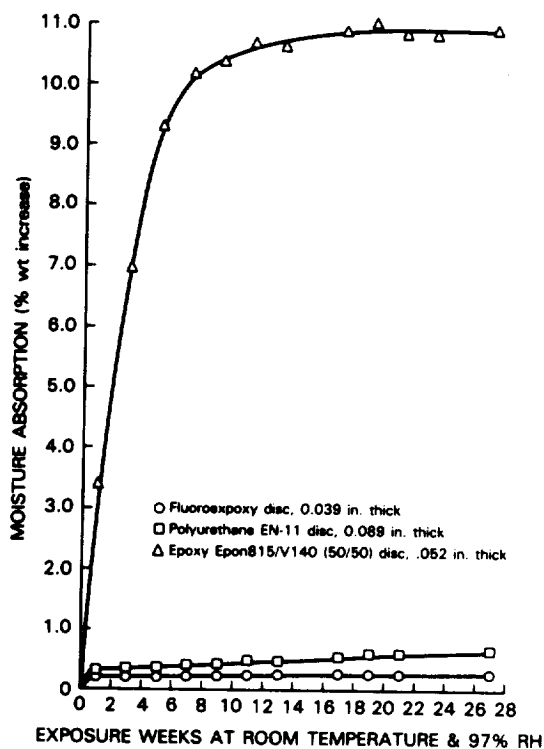


Figure 2. Moisture Absorption  
(Plastic Disk Specimens, 2.284-in. Diameter).

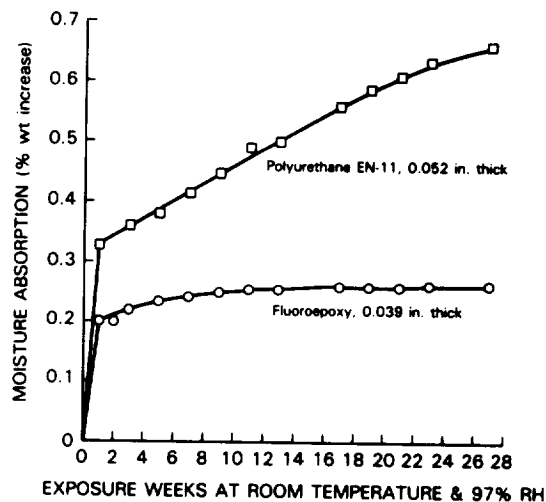


Figure 3. Moisture Absorption  
(Plastic Disk Specimens, 2.284-in. Diameter).

TABLE 1.  
MOISTURE VAPOR TRANSMISSION RATES OF POLYMER FILMS  
(ASTM D1653-72, at 33°C)

Material	F-Content % by wt.	Av. film thickness mm (mil)	Transm. rate mg-mm/(cm <sup>2</sup> -24h)	Remarks
Fluoroepoxy	56	0.045 (1.8)	0.101	Spray coating
DuPont Teflon PFA	72	0.045 (1.8)	0.018	290°C softening
FEP	76	0.045 (1.8)	0.011	240°C softening
Solithane 113/300	0	0.28 (11)	1.19, 1.28	RT 7 days cure
Uralane 5750LV	0	0.34 (13)	1.18	65°C 10 h cure
Uralane 5753LV	0	0.18 (7)	1.21	55°C 24 h cure
Conathane EN11	0	0.32 (13)	1.33	45°C 48 h cure

Note: Fluoroepoxy compound was formulated with I-C8F and adduct amine VI.  
The last four materials are popular polyurethane coatings.

TABLE 2. TENSILE STRENGTH OF ADHESIVES  
(ASTM D2095)

Rod adherend	Fluoroepoxy I C8F/VI		Epoxy Epon 828/V140	
	Strength, psi	Failure mode	Strength, psi	Failure mode
Aluminum	3470±330	Adhesive	7030±258	Adhesive/cohesive
Teflon	640 ±48	Adhesive	330 ±16	Adhesive
Teflon, etched			912±180	(a)

(a) All failed by the separation of the etched surface layer. Strength varied from 710 to 1120 psi. Teflon PTFE was used.

TABLE 3  
ADHESIVE TENSILE STRENGTH OF FLUOROPOXY COMPOUNDS  
IN BONDING TEFLON PTFE WITHOUT ANY SURFACE TREATMENT  
(ASTM 2095)

Resin/Adhesive	F-content in the compound	Tensile strength psi
Fluoroepoxy I-C8F	56.4%	550 ± 46
I-C6F	54.6%	601 ± 87
I-C6F/COF	50.2%	567 ± 163
I-C6F/COF	46.9%	610 ± 55
I-C6F/COF	46.7%	590 ± 78
I-COF	45.9%	358 ± 51
Epon828/V140 (50/50)	---	330 ± 16

Note: All the fluoroepoxy compounds were formulated stoichiometrically with adduct amine VI as the curing agent, and cured at 80°C for 2 hours after room temperature gelation. The F-content of compounds I-C6F/COF was varied by changing the ratio of C6F to COF.

TABLE 4  
ANTIFOAMING EFFECT OF FLUORODIEPOXIDES

Epoxy compound	Fluoroepoxy content by wt.	Number of foam-up in vac. deaeration
Epon828/V140 (50/50)	----	>25
Epon828/V140/C8F (4)	1.09%	1/2
(5)	0.55%	1
(6)	0.23%	4
(7)	0.10%	4
Epon828/TETA (14 phr)	----	13
Epon828/TETA/C8F (8)	0.23%	1
(9)	0.096%	1
Thermoset 321	----	22
Thermoset 321, +C8F	0.19%	1/2
Scotchweld 2216	----	34
Scotchweld 2216, +C8F	0.14%	2
Epon 828/TETA/C6F	0.23%	1/2

Notes: 1. Containers used in the experiments are PE cups, 180 ml size, 6 cm depth. Epoxy compounds weighed about 30 gms and took 2 cm or less in depth in the cup. The number of foam-up means the cycle number of foaming up to ther cup top and breaking it down.  
2. V140 is Versamid 140. TETA is triethylenetetramine. Thermoset 321 and Scotchweld 2216 are the products of Thermoset Plastics and 3M Company respectively.  
3. Fluorine content of the fluorodiepoxides: C8F 58.6%; C6F 56.5%.

TABLE 5  
ADHESIVE LAP SHEAR TENSILE STRENGTH  
(ASTM D1002)

Epoxy compound	Fluoroepoxy content by wt.	Lap shear strength psi
Epon828/V140 (50/50)	---	2390 ± 211
Epon828/V140/C8F	0.10%	2293 ± 297
	0.50%	2599 ± 297
Scotchweld 2216	----	3186 ± 301
Scotchweld 2216, +C8F	0.20%	3036 ± 169

Note: Scotchweld 2216 is a product of 3M Company.



## DUAL ION BEAM PROCESSED DIAMONDLIKE FILMS FOR INDUSTRIAL APPLICATIONS

M.J. Mirtich

National Aeronautics and Space Administration  
Lewis Research Center  
Cleveland, Ohio 44135

M.T. Kussmaul

Sverdrup Technology, Inc.  
Lewis Research Center Group  
Brook Park, Ohio 44142

and

B.A. Banks and J.S. Sovey

National Aeronautics and Space Administration  
Lewis Research Center  
Cleveland, Ohio 44135

## INTRODUCTION

Diamondlike carbon (DLC) films have generated interest due to the unique properties of the material. The durable, smooth, adherent films are quite transparent and impervious to reagents which dissolve graphitic and polymeric carbon structures. These amorphous films have desirable characteristics similar to those of diamond, such as extreme hardness, high transparency (both at visible and infrared wavelengths), high electrical resistivity, low coefficient of friction, and chemical inertness. They do not, however, have the long range order of the diamond crystal structure. Most importantly, DLC films can be deposited on surfaces at room temperature, allowing a broad range of applications. At NASA Lewis Research Center, single and dual beam ion source systems are used to generate amorphous DLC films, which have been evaluated for a variety of applications including protective coatings on transmitting materials, power electronics as insulated gates and corrosion resistant barriers. A list of the desirable properties of DLC films along with potential applications identified at NASA Lewis and those focused DLC film applications presently being developed at Diamonex, Inc. are presented herein.

## DUAL BEAM ION SOURCE SYSTEM AND DLC DEPOSITION PROCEDURE

A 30-cm diameter ion source with its extraction grids masked to 10 cm in diameter is used to directly deposit DLC films. The ion source, developed for electric propulsion technology, uses argon gas in the hollow cathode located in the main discharge chamber, as well as in the neutralizer (ref. 1). After a discharge is established between the cathode and the anode, methane ( $\text{CH}_4$ ) is introduced through a manifold into the discharge chamber. For the depositions presented in this paper the molar ratio of  $\text{CH}_4$  to argon was 0.28. This ratio was found to be ideal for generating films at NASA Lewis. In these experiments the total ion beam energy is the sum of the discharge voltage and the screen grid voltage and is around 100 eV. Typically current densities at these conditions are  $1 \text{ ma/cm}^2$  at a distance of 2.5 cm axially downstream of the grids (ref. 2). Films are deposited at these conditions on Si and  $\text{SiO}_2$  at deposition rates as high as  $71 \text{ \AA/min}$  to film thicknesses as great as  $1.5 \text{ }\mu\text{m}$ .

It is believed that the amorphous carbon films are produced under conditions where both growth and sputtering occur simultaneously, increased sputtering may decrease the number of graphite precursors incorporated in the films and hence improve film quality.

In addition, Marinow and Dobrew (ref. 3) have found that active sites for nucleation are created, and the growth and coalescence of the nuclei enhanced due to an increased mobility of the condensing atoms when film structures are bombarded by inert gas beams. With these factors in mind, a dual beam system was created by adding an 8-cm diameter argon ion source. This system, shown in figure 1, was used to generate another set of DLC films. The 8-cm source, using a filament cathode, was located at a  $12^\circ$  angle with respect to the 30-cm source and 25 cm from the substrate. There was no observed interaction between the two sources or the ion beams during operation.

The 8-cm ion source was used to direct a beam of energetic argon ions (200 to 600 eV) at a current density of  $25 \mu\text{a}/\text{cm}^2$  on the substrates while the deposition from the 30-cm ion source was taking place. The beams were approximately monoenergetic, however no mass selection was attempted to determine species.

## PROPERTIES OF DIAMONDLIKE FILMS

### OPTICAL PROPERTIES OF FILMS

Shown in figure 2 is the spectral transmittance for DLC films generated using the single and dual beam ion source systems. The 1500 Å thick dual beam film has greater transmittance at all wavelengths when compared to the 1500 Å thick single beam films. Some thinner DLC films (800 to 1500 Å thick) look clear to yellowlike in appearance and films thicker than 1500 Å usually look brown. A DLC conformed complete cover coating can usually be obtained for films as thin as 500 Å. The 500 Å thick film presented in figure 2 has transmittance values greater than 90 percent at wavelengths greater than 7000 Å. It is the goal of a joint effort by NASA Lewis and Diamonex, Inc. to obtain clear DLC films, 1000 Å thick, with a transmittance of 85 percent at a wavelength of 5000 Å (the peak of the visible spectrum). Diamonex, Inc. (90 Winsor Drive, Allentown, Pa. 18106) has been in a DLC film technology transfer program with NASA Lewis.

Figure 3 shows the results of some of the on-going studies at NASA Lewis using the dual beam system to reduce the absorption at 5000 Å and hence improve the transmittance. Shown in the figure is the transmittance at  $0.5 \mu\text{m}$  (the peak of the visible) for a DLC film on quartz for various dual beam gaseous deposition conditions. Figure 3 clearly shows that a higher transmittance can be obtained by using the dual beam system where the second ion source uses hydrogen gas. The role of energetic hydrogen in improving transmittance of the DLC film is still under investigation.

The infrared transmittance of DLC films has been shown to be 100 percent transmitting in the wavelength region between 2.5 and  $20 \mu\text{m}$ . DLC films of controlled thickness have been shown to improve the transmittance of Ge infrared optics (refs. 4 and 5). Because the index of refraction ( $\sim 2$ ) of DLC films is intermediate between Ge and air, a DLC coating is capable of improving overall transmittance by reducing interface reflection losses. A DLC coating can also be used to reduce reflection losses of Si. The high hardness of these films also imparts abrasion protection to these infrared optical materials.

Listed in table I are values of H/C ratio, optical band gap, index of refraction, absorption coefficient and solar transmittance for DLC films generated using the single or dual beam ion sources. Certain properties are enhanced by using a dual beam system, mainly the solar transmittance which is higher, as is the index of refraction, and a lower absorption coefficient, when compared to the single ion source films. The optical band gap and density for the single and dual beam system films are similar. Also listed in table I are values of resistivity, density, and adherence. Since all of the NASA Lewis DLC films were processed in oil diffusion-pumped facilities without the benefit of liquid nitrogen cryo-traps, even higher quality films might be obtained by using a helium cryopumped facility.

## CHEMICAL AND PHYSICAL PROPERTIES

The DLC single and dual beam films were subjected to a solution of 3 parts  $\text{H}_2\text{SO}_4$  and 1 part  $\text{HNO}_3$  (concentrated acids, by volume) at  $80^\circ\text{C}$  for periods up to 20 hr. The films on silicon were unaffected by the reagent. These results clearly indicate that the films are far more resistant to chemical etching than normal polymeric hydrocarbons or graphite. Dr. J.C. Angus (ref. 6) measured the diffusion coefficient of Argon, trapped at the DLC film/substrate interface, to be less than  $10^{-18} \text{ cm}^2/\text{sec}$ . This value is 10 orders of magnitude (10 billion) times less than conventional hydrocarbons or polymers. Thus DLC films are one of the best known hermetic barriers. This suggests the potential use of the DLC films as chemical and/or a diffusion barrier for microelectronic or optical components.

Transmission electron microscopy at 30 000 X showed the films to be smooth and essentially free of features. No pinholes or other defects were observed.

The adherence of the films on quartz was measured following the procedure used by Mirtich (ref. 7). The adherence of the films generated with either the single or dual beam systems were as good as the maximum adherence of the Sebastian Adherence Tester<sup>®</sup> used in the measurement ( $\sim 5.5 \times 10^7 \text{ N/m}^2$  or 8000 psi). The film adherence was extremely high and frequently found to exceed the cohesion of the substrate such that portions of quartz gave way with the film still intact.

Some films deposited with the dual beam on Si have been stored for 10 years and show no visible signs of deterioration. Figure 4 shows a plot of the coefficient of friction for a diamondlike carbon film sliding on steel as a function of relative humidity. For a humidity less than one percent the coefficient of friction is very low, indicating a potential use of a DLC film as a wear resistant surface for inert or vacuum environments.

Scratch tests were performed on DLC films deposited on quartz by rubbing  $\text{SiO}_2$  particles ( $\sim 80 \mu\text{m}$ ) over the surface. Figure 5 shows the result of these tests. The DLC film protected the quartz from the abrasive effects of the  $\text{SiO}_2$  particles.

Table II lists the properties of diamondlike films and compares them to those of natural diamond and polycrystalline diamond films (ref. 8). In a couple of categories DLC films do not quite measure up to those of polycrystalline or natural diamond. For instance, natural diamond and polycrystalline diamond have thermal conductivities of 900 and 700 W/mK respectively; whereas, a DLC film is only 7 W/mK. Likewise, the resistivity of DLC films is not as high as that of diamond. Although the microhardness of DLC films ( $5\,000 \text{ kg/mN}^2$ ) is approximately half ( $1/2$ ) of that of diamond, it is twice as hard as hard corundum ( $2\,085 \text{ kg/mN}^2$ ) and harder than sapphire. This hardness value makes DLC films a good coating when a hard surface is desirable (i.e., magnetic memory protection, eyeglass lenses, etc.). The coefficient of friction (0.1) for a diamondlike film at low humidity is similar to that of natural diamond. Thus any material that is affected by wear has the potential to benefit from the reduced friction properties and hardness of diamondlike coatings.

The distinguishing difference between DLC film and polycrystalline film is that DLC films are smooth, adherent and are deposited at room temperature or less. This feature of DLC films alone makes them a popular coating choice in a vast area of applications. Whereas, polycrystalline diamond films, which are rough, not very adherent and not specularly transmitting as yet, have the limiting feature of having the substrate at temperatures greater than  $600^\circ\text{C}$  during deposition. This limits polycrystalline diamond film's applications.

## POTENTIAL APPLICATIONS

A feasibility study of commercial applications of diamondlike films was done at the NASA Lewis. The objective of the study was to evaluate, rank, recommend and plan commercial applications for NASA-developed diamondlike film deposition technology. The approach was to

identify potential users, prioritize potential applications and assess the feasibility of selected applications. A questionnaire was mailed to 230 companies. The number of responses returned was 39 with a total of 44 applications identified. Table III presents the top 12 potential applications identified by the study that appear, at the moment, feasible using diamondlike films. There are, of course, many other applications not listed here but bear mentioning like durable luster enhancing coatings for gemstones, abrasion resistant coatings for plastics, etc.

Shown in table IV are the focused diamondlike film applications at Diamonex, Inc., is a new corporation founded in 1990 whose principle technologies are polycrystalline diamond and diamondlike films. They have purchased an exclusive license to the NASA Lewis dual ion beam patent to make diamondlike films. Diamonex has the rights to also sublicense the patent. Because of some of the properties of DLC films listed in table II, the technical feasibility and the user need, they have focused their resources on making abrasion-resistant optical coatings their number one priority. Other applications listed include the use of DLC films as a hermetic seal of transparent substrates and wear protection of nonoptical substrates.

### SUMMARY

A major advantage of DLC films is that they are smooth and adherent and can be deposited on substrates at room temperature. These features allow DLC films to be used readily in many applications where properties of DLC films such as hardness, scratch resistance, hermeticity and transparency are important. Included in this paper are prioritized lists of many potential DLC film applications.

Diamonex, Inc., which has a license to NASA Lewis dual ion beam DLC film patent, has estimated the DLC film potential market to be several billion dollars. The abrasion-resistant optical coatings market for eyeglass lenses alone is estimated at a half a billion dollars. Diamonex is authorized to grant sub-licenses under their exclusive license arrangement from NASA.

### REFERENCES

1. Sovey, J.S.: AIAA paper 76-1017, Key Biscayne, Fla., Nov., 1976.
2. Mirtich, M.J.: AIAA paper 81-0672, Las Vegas, Nev., April, 1981.
3. Marinow, M., and Dobrew, D.: Thin Solid Films, Vol. 42, 1977.
4. Pellicore, S.F., Peterson, C.M., and Henson, T.P.: J. Vac Sci. Technology, A4, 2350 (1986).
5. Koidl, P., Bubenzer, A., Dischler, B., Pro. Soc. Photo-opt. Instrum. Eng. 381, 186 (1983).
6. Angus, J.C., Mirtich, M.J., and Wintucky, E.G.: Metastable Materials Formation by Ion Implantation, Elsevier Science Publishing Co., Inc., 1982.
7. Mirtich, M.J.: J. Vac. Sci. and Tech., 18, 2, March, 1981.
8. Kempfer, L.: Diamond: A Gem of a Coating. M. E. May, 1990, pp. 26-29.

**TABLE I - OPTICAL, CHEMICAL AND PHYSICAL PROPERTIES  
OF DLC FILMS GENERATED USING CH<sub>4</sub> WITH  
SINGLE OR DUAL BEAM ION SOURCES**

	Single beam CH <sub>4</sub> /A = 28 percent	Dual beam CH <sub>4</sub> /A = 28 percent
H/C ratio	--	1.0
Resistivity, Ω-cm	8.66x10 <sup>6</sup>	3.35x10 <sup>6</sup>
Optical band gap (eV)	0.382	0.343
Density, gm/cm <sup>3</sup>	1.8	1.8
Index of refraction	2.0	2.46
Absorption coefficient/cm at 5000Å	5.15x10 <sup>4</sup>	4.26x10 <sup>4</sup>
Solar transmittance $t = \frac{\int Q(\lambda)T(\lambda)d(\lambda)}{\int Q(\lambda)d\lambda}$	0.519 Film = (1500 Å thick)	0.648 (1500 Å)
Adherence (Quartz)	>5.5x10 <sup>7</sup> N/m <sup>2</sup> >8000 psi	>5.5x10 <sup>7</sup> N/m <sup>2</sup> >8000 psi

Where: Q(λ) = 6000°K black body  
T(λ) = transmittance at a given wavelength

## TABLE II

Properties of Diamondlike Films Compared to Natural Diamond and Polycrystalline Diamond Films

Properties	Diamond	Polycrystalline Diamond Films	Diamond-Like Carbon Films
Microhardness (kg/mm <sup>2</sup> )	10,000	8-10,000	3-5,000
Resistivity (ohm cm)	$10^{13} - 10^{16}$	$10^{10} - 10^{15}$	$10^7$
Thermal Conductivity (W/mK)	2,000	700	(7) Low
Refractive Index	2.4	2.3 - 2.4	2.4
Density (g/cm <sup>3</sup> )	3.5	3.2 - 3.4	1.8
Friction Coefficient	0.1	--	<0.1 (at low humidity)
Chemical Resistance	High	High	High
Visible and Infrared Transparency	High	High	High
Visible and Infrared Specular Transparency	High	Low	High
Adherence	--	Low	High
Substrate Temperature During Deposition	--	>600° C	Room temperature or lower
Film Roughness	--	Rough	Smooth

**TABLE III**  
**Diamondlike Film Potential Applications Identified at LeRC**

DLC Film Application	Desirable DLC Film Properties
1. Protective coating for sunglass lenses	Hardness, scratch resistance
2. Protective coating for eyeglass lenses	Hardness, scratch resistance, transmittance
3. Hermetic coating for eyewear	Hermeticity, hardness, scratch resistance, transmittance
4. Abrasion, moisture resistant coating for optical surfaces (visible and infrared)	Hermeticity, hardness, scratch resistance, transmittance
5. Magnetic recording head	Hermeticity, hardness, scratch resistance
6. Coating computer hard disk	Hermeticity, hardness, scratch resistance
7. Abrasion resistant coating for optical windows in bar code scanners	Hardness, scratch resistance
8. Biomedical applications	Biocompatibility, hermeticity
9. Chemically resistant protective coating	Hermeticity, hardness
10. Enhanced IR transmittance of Ge and Si infrared optics	IR transmittance, index of refraction, abrasion protection
11. Cutting blades	Smoothness, hardness, hermeticity
12. Abrasion resistant non-stick coating for cookware	Hardness, scratch resistance

## **TABLE IV**

### **Focused Diamondlike Film Applications at Diamonex, Inc.**

**(1) Abrasion-resistant optical coatings**

- Plastic sunglass lenses
- Plastic eyeglass lenses
- Other optical substrates such as glass

**(2) Chemical/environmental protection ("hermetic sealing") of transparent substrates**

- Quartz
- Glass
- Plastics

**(3) Wear protection of non-optical substrates**

- Magnetic disks
- Cutting surfaces
- Other wear parts



ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 1.—Dual beam ion source for deposition of films with diamondlike properties.

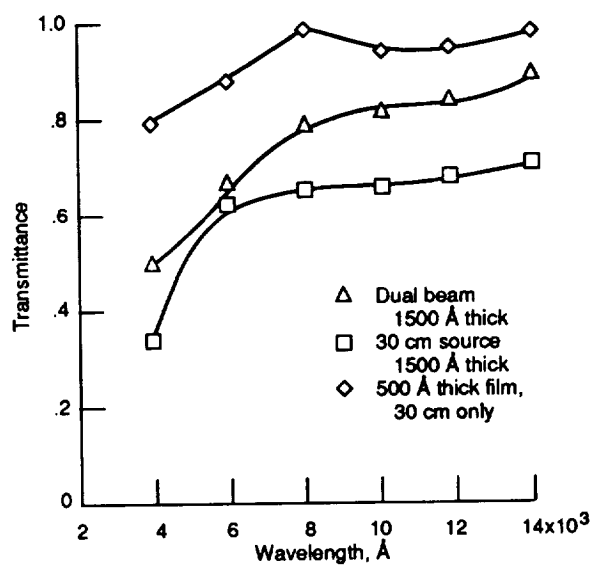


Figure 2.—Transmittance versus wavelength for DLC films using  $\text{CH}_4$  in dual beam or single ion sources.

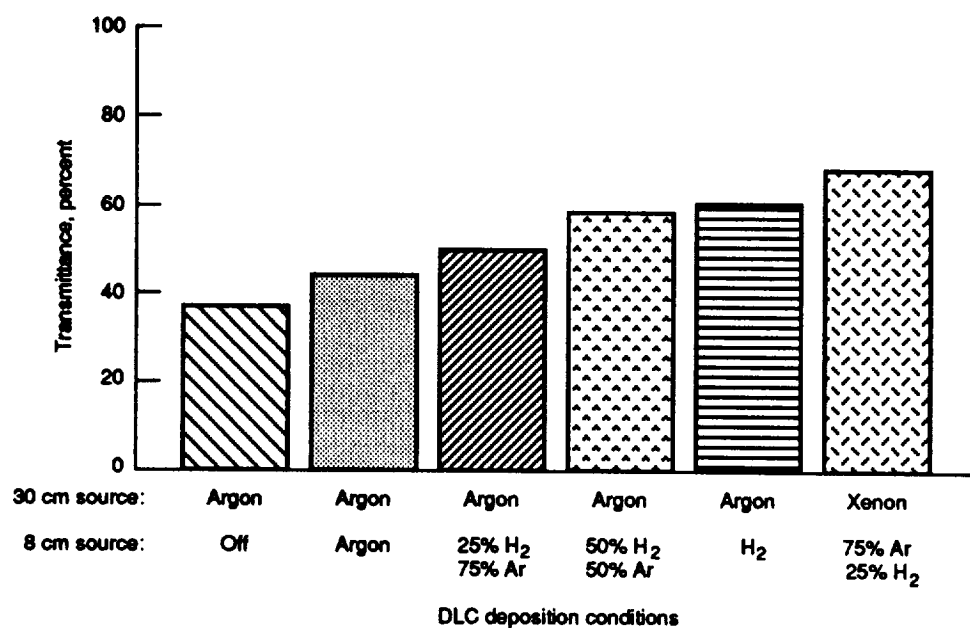


Figure 3.—Transmittance at 0.5  $\mu\text{m}$  for 1000 Å thick DLC films deposited on quartz under various dual beam conditions.

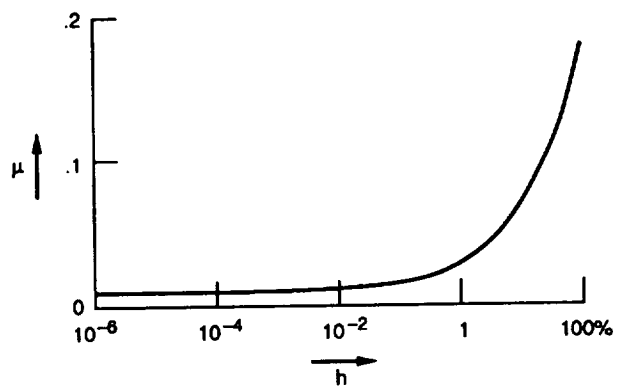


Figure 4.—Coefficient of friction  $\mu$  for a diamondlike carbon film sliding on steel, as a function of the relative humidity  $h$ .

ORIGINAL PAGE IS  
OF POOR QUALITY

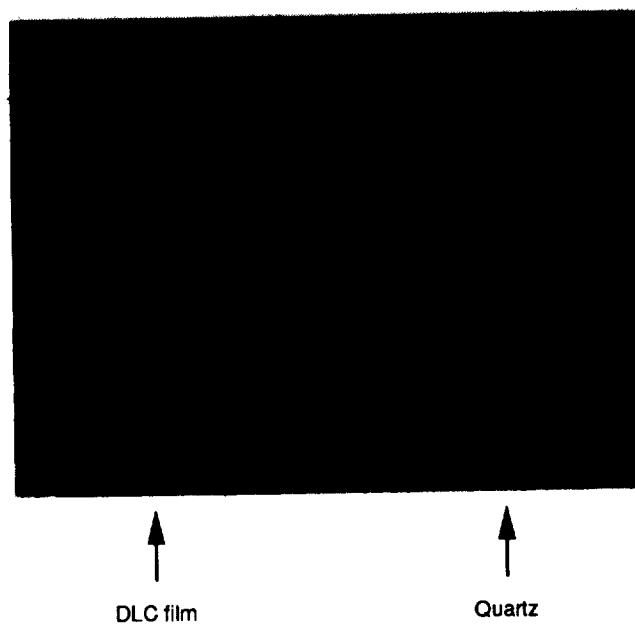


Figure 5.— Scratch test of DLC film using  $\text{SiO}_2$  particles.

## PLASMA-POLYMERIZED COATING FOR POLYCARBONATE: SINGLE-LAYER, ABRASION RESISTANT, AND ANTIREFLECTION

Theodore Wydeven  
NASA Ames Research Center  
Moffett Field, California 94035

### ABSTRACT

Plasma-polymerized vinyltrimethoxy silane films were deposited on transparent polycarbonate substrates. The adherent, clear films protected the substrates from abrasion and also served as antireflection coatings. Post-treatment of the vinyltrimethoxy silane films in an oxygen glow discharge further improved their abrasion resistance. The coatings were characterized by elemental analysis of the bulk, ESCA analysis of the surface, transmission, thickness, abrasion resistance, haze, and adhesion. This patented process is currently used by the world's largest manufacturer of non-prescription sunglasses to protect the plastic glasses from scratching and thereby increase their useful lifetime.

### INTRODUCTION

Traditionally, ground and polished tempered glass was the material of choice for manufacturing sunglass lenses. However, in recent years the many virtues of polarized and other plastic lenses have come into sharp focus. The advantages of plastic lenses include excellent optics, lightweight and therefore more comfortable than glass, easy to shape, reasonable to make, vastly better absorption of harmful ultraviolet radiation, and resistance to shattering. The major disadvantage of most plastic lenses is susceptibility to scratching. Uncoated plastic lenses develop a haze after a relatively short number of wearings which greatly reduces visibility.

There are many other commercial applications for transparent plastics. Two large industries, automobile and construction, have numerous needs for transparent plastics. Automobile lenses and dashboard covers and glazing applications in construction are just some of the needs of these industries for transparent plastics. As in the optics industry, resistance of the plastic to scratching and abrasion is desirable for most automobile and construction applications.

Space helmet visors are also made of plastic and it is this application that led NASA to develop a transparent scratch resistant coating for plastics. Polycarbonate, an impact resistant and transparent plastic is generally used for space helmets. However, polycarbonate also suffers from the disadvantage of most plastics for optics applications, i.e., it is highly susceptible to scratching. NASA therefore embarked on a program to develop a scratch resistant coating for polycarbonate space helmet visors. The coating process chosen for this development effort was plasma polymerization. This process was chosen for several reasons;

- a) deposition occurs near ambient temperature, making the process useful for coating temperature sensitive plastics.
- b) plasma polymerized coatings are conformal and uniform in thickness.
- c) practically any organic vapor can be used as starting material for depositing coatings by plasma polymerization; this feature allows one to deposit coatings covering a broad range of composition and properties.
- d) it can easily be scaled for coating large samples.
- e) multiple treatments and processes can be carried out in the same plasma reactor.

The starting material chosen for preparing a scratch resistant coating was based primarily on its chemical composition and structure. The composition was chosen so that the deposited film would mimic in some respects the composition of glass (a material rich in silicon and oxygen) which is known to be much more scratch resistant than plastic. The structure was chosen so that reasonable deposition rates

would be achieved thus minimizing the time required to deposit a coating of sufficient thickness to protect the underlying plastic from being scratched. Enhanced deposition rates can generally be achieved by plasma polymerization if the starting material contains a vinyl group or carbon-carbon double bond. With these requirements in mind the starting material chosen was vinyltrimethoxy silane  $[(\text{H}_3\text{CO})_3\text{-Si-CHCH}_2]$ . Since coatings are prepared from gases or vapors by plasma polymerization, the vapor pressure of the starting material was also considered in the selection process. Vinyltrimethoxy silane is a liquid with sufficient vapor pressure at room temperature to be a useful compound for plasma polymerization.

This paper describes the preparation and properties of a transparent scratch resistant and antireflection coating for polycarbonate. A two-step process was used for preparing the coating. The first step involved deposition of a plasma polymerized coating of vinyltrimethoxy silane on the polycarbonate surface and the second step involved hardening of the coating with an oxygen plasma. The second step of the process has been granted US Patent 4,137,365.

## EXPERIMENTAL

### *Deposition Chamber and Materials*

The deposition chamber shown in Figure 1 was used to deposit and harden the transparent scratch resistant coating. Copper sheet electrodes were wrapped around the outside of the reactor and were attached to a 13.56 MHz radiofrequency power supply through a manually controlled impedance matching network. The transparent plastic substrates used for the coating experiments were cut from 0.64 cm thick polycarbonate sheet stock manufactured by Rohm & Haas and sold under the tradename Zelux-C. The square substrates had an exposed surface area of 6.45 cm<sup>2</sup> during deposition and were held in a plastic frame supported on a glass plate located inside the reactor as shown in the figure. The frame ensured uniformity in thickness of the deposited coatings from the center to the edge of the samples. The substrates were cleaned before being coated by degreasing them in vapors of trichlorotrifluoroethane. The plasma polymerized coatings were prepared from vinyltrimethoxy silane vapor that was delivered to the deposition chamber from a degassed thermostatted reservoir of the liquid. The coatings were further hardened in the glow of an oxygen plasma. The oxygen for these experiments was obtained from a cylinder of oxygen having a reported purity of 99.5%.

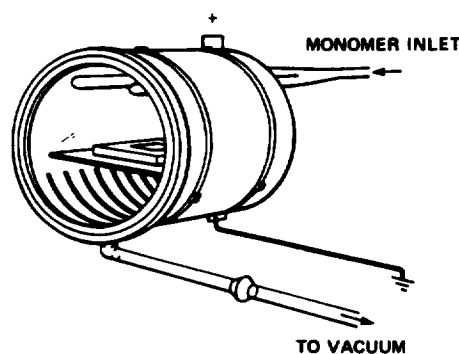


Figure 1. Chamber for depositing and hardening coatings using a low temperature plasma.

### *Deposition and Film Hardening Procedure*

The reactor, with substrate in position, was initially evacuated to a background pressure of 0.7 N/m<sup>2</sup> (0.005 torr) or less before initiating flow of vinyltrimethoxy silane vapor. The flow rate of the monomer was adjusted to 3.2x10<sup>-9</sup>m<sup>3</sup> (STP)/sec (0.19 cm<sup>3</sup>(STP)/min), which yielded a pressure of 2.8 N/m<sup>2</sup> (0.021 torr), before striking a discharge. A pulsed plasma mode, instead of a continuous wave (CW) mode, was used for film deposition. The pulsed mode reduces substrate heating during deposition. The peak-to-peak voltage during the on-time of the pulse was 400 V and the peak-to-peak current to ground was 0.6 A. The pulse time was 1.5 msec on and 1.5 msec off. At the end of a pre-determined deposition time,

monomer flow to the deposition chamber was shutoff and residual monomer was pumped from the chamber. The hardening of the deposited coating was then begun without removing the coated substrate from the reactor. The following conditions were used for hardening the siloxane coating with an oxygen plasma:

Mode of plasma operation:	CW
Peak-to-peak voltage:	390 V
Peak-to-peak current:	0.58 A
Oxygen flow rate:	$2.9 \times 10^{-8} \text{ m}^3(\text{STP})/\text{sec}$ ( $1.7 \text{ cm}^3(\text{STP})/\text{min}$ )
Initial oxygen pressure (discharge off):	$13 \text{ N/m}^2$ (0.098 torr)
Post-treatment duration:	600 sec

Following the hardening step, substrates were removed from the reactor for evaluation of the coatings.

#### *Analytical Procedures for Evaluating the Coatings*

The test equipment and procedures followed for determining the elemental analysis, x-ray photoelectron spectra (ESCA), transmission, abrasion resistance, haze, adhesion, and thickness of the siloxane coatings on polycarbonate can be found elsewhere (1,2).

### RESULTS AND DISCUSSION

Samples for elemental analysis of plasma polymerized vinyltrimethoxy silane films were obtained by scraping the coating from a glass plate which was coated with the polymer. The average weight percent of carbon, hydrogen and silicon dioxide in plasma polymerized vinyltrimethoxy silane films (without oxygen post-treatment) derived from elemental analysis by a combustion method of three samples was:

Element	Weight Percent
C	35.84
H	6.64
SiO <sub>2</sub>	48.56
Si	22.69

Since this analysis was done by a combustion method, the amount of oxygen in the samples could not be measured. The amount of silicon in the samples was calculated from the amount of silicon dioxide. The stoichiometric composition of vinyltrimethoxy silane is 40.52, 8.16, and 18.95 weight percent carbon, hydrogen, and silicon, respectively. In comparing the elemental analysis of the bulk plasma polymer with the starting monomer composition it is apparent that plasma polymerized vinyltrimethoxy silane films are deficient in carbon and hydrogen and rich in silicon relative to the starting monomer. A deficiency in carbon and hydrogen could have arisen if a carbon/hydrogen entity, such as a methyl group, was stripped from the starting monomer during plasma deposition, leaving a film with an enhanced amount of silicon but deficient in carbon and hydrogen.

X-ray photoelectron spectral analysis was chosen to determine the elemental composition of the *surfaces* of untreated and oxygen plasma post-treated vinyltrimethoxy silane plasma polymerized films. The results derived from these analyses were:

Element	Untreated film, atom %	O <sub>2</sub> plasma treated, atom %
C	21.7	13.7
O	12.1	24.2
Si	13.3	9.21

Since ESCA cannot be used to analyze for hydrogen, it was assumed that the hydrogen concentration on the surface was the same (52.87 atom %) as in the bulk plasma polymer before calculating surface composition from the ESCA data. The atom % of the elements carbon, oxygen, and silicon in stoichiometric vinyltrimethoxy silane monomer are 23.81, 14.28, and 4.76, respectively. It is clearly

evident from the ESCA results that oxygen post-treatment of the plasma polymerized vinyltrimethoxy silane coating enriched the surface in oxygen at the expense of carbon, possibly by volatilizing the carbon as carbon monoxide or carbon dioxide. The oxygen plasma treated surface is also greatly enriched in oxygen and silicon and deficient in carbon relative to the starting monomer.

The intent in treating plasma polymerized vinyltrimethoxy silane coatings with an oxygen plasma was to change their composition into more of a glass-like material, rich in silicon and oxygen, while maintaining transparency and improving resistance to scratching and abrasion. It is clear from analysis of the treated surface that the composition has been transformed by the oxygen plasma into a more oxygen rich material having less surface carbon.

An image viewed through a lense covered with an antireflection coating appears sharper than the image viewed through an uncoated lense. Hardened plasma polymerized vinyltrimethoxy silane was found to be an antireflection coating for polycarbonate. Figure 2 shows transmission spectra for an uncoated polycarbonate substrate and substrates coated with oxygen post-treated plasma polymerized vinyltrimethoxy silane films. It is apparent from these spectra that the coated samples of polycarbonate transmit more light in the visible region than the uncoated sample. For example, at 5500 Å, the midpoint of the visible region of the spectrum, the transmission of the coated samples is greater by about 5%. In the cases of the coated samples, two different coating thicknesses were deposited, 939 and 2818 Å. These physical thicknesses correspond to optical thicknesses of 1375 and 4125 Å, respectively. The optical thickness,  $nd$ , where  $n$  is the refractive index and  $d$  is the physical thickness of the coating, was calculated using 1.464 for the refractive index of the coating. The refractive index at 5893 Å (the  $\text{Na}_D$  line) was measured with an Abbe refractometer.

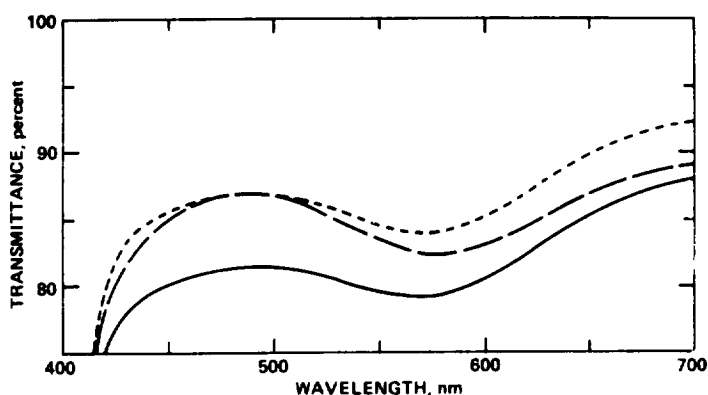


Figure 2. Transmission spectra of uncoated and coated (both sides) polycarbonate. — uncoated, - - - coated with 939 Å of plasma polymerized vinyltrimethoxy silane, and - · - coated with 2818 Å of vinyltrimethoxy silane.

The abrasion resistance of samples of uncoated and coated polycarbonate were determined by measuring haze before and after abrasion with an eraser. The force applied to the eraser when abrading the samples using the apparatus and procedure described previously was  $1.4 \times 10^3 \text{ kg/m}^2$  (1). The following table shows haze values for the various samples studied:

Sample description	% Haze	Notes
Uncoated PC	$5.2 \pm 0.8$	Ave. for 10 samples
*Coated PC	$4.1 \pm 0.5$	Ave. for 3 samples
*Coated & $\text{O}_2$ treated PC	$2.5 \pm 0.2$	Ave for 6 samples
Uncoated & unabraded PC	$2.3 \pm 0.1$	Ave. for 10 samples
*Coated & unabraded PC	$2.2 \pm 0.1$	Ave. for 3 samples

PC = polycarbonate

\*Coating thickness = 2818 Å

From this data it is evident that the least amount of damage from abrasion occurred in those coated samples which were post-treated in an oxygen plasma. The haze for oxygen post-treated samples was only 0.2% greater than uncoated and unabraded polycarbonate. It is also evident from this data that coated and unabraded samples had the same haze as uncoated and unabraded polycarbonate within experimental error. This finding shows that neither the plasma polymerized silane coating nor oxygen post-treatment increased the haze of polycarbonate. Another important conclusion that can be derived from the haze data is that oxygen post-treatment improved the abrasion resistance as evidenced by a reduction in haze of 1.6% when compared with coated polycarbonate without oxygen treatment. It should also be noted that untreated plasma polymerized vinyltrimethoxy silane coatings also increased the abrasion resistance of polycarbonate as indicated by a reduction in haze of 1.1%. The improved abrasion resistance that the coatings developed here provide for a polycarbonate surface may be due to a greater hardness of the coating compared with uncoated polycarbonate, as well as less friction between the coating and the abrading material, in this case the eraser.

Adhesion of the plasma coating to polycarbonate was measured using the cellophane tape test as described in Military Specification MIL-C-675A. The coatings consistently passed this test without any special treatment of the polycarbonate surface other than the cleaning procedure described in the experimental section of this paper.

## CONCLUSIONS

A two step low temperature plasma process has been developed for depositing a transparent and antireflection scratch resistant coating on polycarbonate. The first step of the process involves the deposition of a plasma polymerized coating of a siloxane on a clean polycarbonate surface. The second step is a hardening step and involves the treatment of the siloxane coating with an oxygen plasma. Both steps are carried out in the same reaction chamber.

## REFERENCES

1. Wydeven, T.: Plasma Polymerized Coating for Polycarbonate: Single layer, Abrasion Resistant, and Antireflection. Applied Optics vol 16, no. 3, 1977, pp. 717-721. See also Applied Optics vol 16, no. 5, 1977, p. 1138.
2. Wydeven, T.; and Hollahan, J.: United States Patent 4,137,365: Oxygen Post-treatment of Plastic Surfaces Coated with Plasma Polymerized Silicon Containing Monomers, Jan. 3, 1979.



## THE PM-200 LUBRICATION SYSTEM

Harold E. Sliney  
NASA Lewis Research Center  
Cleveland, OH 44135

## ABSTRACT

Plasma sprayed composite coatings of metal-bonded chromium carbide with additions of silver and thermochemically stable fluorides were previously reported to be lubricative in pin on disk bench tests from room temperature to 900° C. An early coating formulation of this type, designated as PS200, was successfully tested as a cylinder coating in a Stirling engine at a TRRT of 760° C (1450° F) in a hydrogen atmosphere, and as a backup lubricant for gas bearings to 650° C (1250° F). A subsequent optimization program has shown that tribological properties are further improved by increasing the solid lubricant content. The improved coating is designated as PS212. The same powder formulation has been used to make free-standing powder metallurgy (PM212) parts by sintering or hot isostatic pressing. The process is very attractive for making parts that cannot be readily plasma sprayed such as bushings and cylinders that have small bore diameters and/or high length to diameter ratios. The properties of coatings and free-standing parts fabricated from these powders are reviewed.

## INTRODUCTION

This paper updates development and application test results with PS200 and PS212 plasma-sprayed composite coatings that have demonstrated self-lubricating capability from low temperature to 900° C. The formulation and the basic friction and wear properties of these coatings are described in references 1 to 4. The two compositions differ only in the amount of solid lubricant they contain within a metal-bonded chromium carbide matrix. PS200 contains 10 wt % silver and calcium fluoride/barium fluoride eutectic while PS212 contains 15 percent of each. The coatings have almost identical friction and wear properties, but PS212 is currently favored because it exhibits slightly lower friction than PS200. These coatings have been successfully tested as backup lubricants for compliant (foil) gas bearings, as cylinder liner coating material for the Stirling engine, and as valve guide material for process control valves. They can be plasma sprayed onto nickel and cobalt base super alloys and PS212 has been directly sprayed onto PSZ thermal barrier coatings. The duplex coating has the advantage of providing the dual function of providing thermal insulation and lubrication over a very broad temperature spectrum.

Recently, PM212, a powder metallurgy version of this material, has been developed at NASA Lewis Research Center. Free-standing parts have been prepared both by sintering and also by hot isostatic pressing (HIPing). The results of pin on disk studies of sintered PM212 are presented.

## MATERIAL PREPARATION

Plasma-Sprayed PS212

The plasma-spray procedure for PS212 has been described elsewhere (refs. 2-4). The sintering and HIP procedures for the powder metallurgy composites of the same chemical composition as PS212 are summarized below.

A schematic diagram of the procedure used in preparing sintered PM212 is shown in Figure 1. A commercial blend of chromium carbide with a nickel-cobalt alloy binder is mixed with silver and calcium fluoride/barium fluoride eutectic in a Vee blender. The composition by weight percent is: 70 metal bonded chromium carbide-15 silver-15 eutectic. A green compact of the powders can be prepared by cold pressing in a die or by cold isostatic pressing (CIP). Our compacts were prepared by the CIP process. The powder is first poured into a rubber mold which is then evacuated with a vacuum pump. The mold is then sealed

under vacuum, placed in a perforated aluminum support container, and immersed in a liquid pressure chamber. The powders are compacted hydraulically at high pressure. The cold pressed compact is then removed from the mold and sintered to increase the strength of the compact. Sintering is done at high temperature in a dry hydrogen atmosphere. Care must be exercised to maintain a low dew point in the hydrogen atmosphere to avoid the formation of chromium oxide during the sintering process. A back scatter SEM picture of the sintered composite is given in Figure 2. Porosity and the various components of PS212 can all be discerned in this picture. Sintered PM212 contains about 20 percent pore space.

### Hot Isostatically Pressed PM212

As indicated in Figure 3, the blended powder can either be HIPed directly or it can be first compacted by the CIP process, then HIPed. We chose to precompact the blended powder by the CIP process. The green compact is then removed from the rubber mold, wrapped in graphoil, and placed in a steel can. The can is then vacuum sealed, and placed into a pressure vessel. The vessel is filled with high pressure argon gas and heated to the sintering temperature. After a prescribed time at temperature, the gas pressure and temperature are returned to ambient conditions and the can is removed from the chamber. The compacts made by this process are fully dense. The density is  $6.60 \text{ g/cm}^3$ . The density difference of the sintered and HIPed materials are shown in Figure 4. Photomicrographs of the diamond ground surface of sintered PM212 (Figure 5) show the fluoride eutectic distributed throughout the carbide and metal matrix. Photographs taken with different types of illumination are shown to illustrate that the translucent fluorides cannot be differentiated from porosity when vertical illumination is employed, but they are imaged very well under oblique illumination.

The compressive strength of sintered and HIPed PM212 to  $900^\circ \text{C}$  are shown in Figure 6. The HIPed version is about three times stronger than the sintered version, but both materials retain strength to at least  $700^\circ \text{C}$  that exceeds or equals the room temperature strength of many grades of bronze and carbon sliding contact bearing materials.

### Machining

Depending upon the precision to which the sintered parts or isostatically pressed parts can be processed during their preparation, either very little or quite extensive machining may be required. Parts that are cold mechanically pressed in a precision die before sintering may require no more than a mild deformation sizing step to achieve accurate dimensions. Our parts, that were made from sintered or HIPed blanks, required extensive machining. The most successful process that we found for these materials consists of electrical discharge machining (EDM), followed by diamond grinding of surfaces that are to be the sliding contact bearing surfaces. Diamond grinding was the best finishing method found to insure that the softer phases in the composite are not selectively removed during the finishing process. Some machined PM212 parts are shown in Figure 7.

## **FRICTION AND WEAR RESULTS**

Comparative tests of the friction and wear properties of the sintered composite and the plasma-sprayed coating were performed (ref. 5). In the coating experiments, Rene 41 pins with a 4.76-mm radius tip were slid against Inconel 718 disks coated with PS212 that had been diamond ground to a thickness of 0.25 mm. In tests of the sintered material, PM212 pins were slid against uncoated Rene 41 disks.

The friction coefficients of the composite/metal couples are similar to those obtained with metal/coating couples. This is illustrated in Figure 8 where the friction/temperature characteristics of the two combinations are compared. The experiments were conducted in air, with a 4.9-N load at 1000 rpm (2.7 m/s). The friction coefficient was very stable within a range of 0.3 to 0.35 from room temperature to  $850^\circ \text{C}$ . Wear factors for the composite pins and Rene 41 disks are given in Table I.

The effect of load on friction followed Amonton's law from 4.9 to 29.4 N, i.e.: The friction force

increased proportionately to the applied load. The wear factors also did not change significantly within that load range. The contact pressure for a 29.4 N load and a typical wear scar diameter of 1.3 mm is 22 MPa or about 3000 psi. Therefore, although the dynamic load capacity of PS212 was determined, it was higher than 22 MPa at a sliding velocity of 2.7 m/s.

The effect of sliding velocity on the friction coefficient of sintered PM212 is given in Figure 9. At 760 and 900° C, the friction coefficient decreases with increasing sliding velocity and is 0.2 at the highest velocity of 8.1 m/s. Friction behavior is less straightforward at lower temperatures, but is typically 0.3 to 0.35 except at the lowest sliding velocity at room temperature. At the other velocity extreme, the trend of decreasing friction with increasing sliding velocity suggests that friction coefficients considerably lower than 0.2 may be expected at sliding velocities typical of higher speed shaft seals in turbomachinery.

The influence of sliding velocity on reducing friction is probably due to the increased localized frictional heating at high sliding velocities. This localized heating can be expected to soften the silver and fluoride surface films, thus reducing their shear strength without appreciably reducing the hardness and flow pressure of the material below the near-surface region.

It was observed that the surface finish of diamond ground PS212 improved during sliding. Because of surface porosity of the coatings, even finish ground surfaces had a rms surface finish of about 0.8  $\mu\text{m}$  initially but this was reduced typically to about 0.2 after sliding. The same phenomenon was observed with sintered PM212. Figure 10 shows that this was caused by closing of the surface pores and densification of the material under the sliding contact.

## TECHNOLOGY TRANSFER

There has been considerable interest in the possible use of PS200 and PS212 in industrial applications. These coatings have been successfully tested for a number of applications such as:

- o A backup lubricant for gas bearings
- o A valve system coating for gas flow control valves
- o A cylinder liner coating for an automotive Stirling engine
- o A low friction material for gas turbine, high-speed shaft seals
- o A combustion chamber liner for rotary engines

A gas bearing journal and thrust ring coated with PS212 are shown in Figure 11. PS212 journal coatings have lubricated foil gas bearings during start/stop tests at temperatures up to 700° C (ref. 4). The bearings, which are made of 0.013 cm thick Inconel X-750 were still in good condition after 20,000 start/stops over a programmed repetitive temperature cycle from room temperature to 700° C.

We have previously reported promising results with PS200 coatings on Stirling engine cylinder walls (ref. 6). An engine of the design shown in Figure 12 was tested for over 20 hours with Stellite 6B piston rings sliding against the coating in a hydrogen atmosphere at a top ring reversal cylinder wall temperature of 700° C.

As previously stated, the friction coefficients of these composite materials tend to decrease with increasing sliding velocity. Therefore, low friction and wear can be expected during high speed, high temperature operation of PS200 or PS212 coated components such as the turbine shaft seal shown in Figure 13.

## CONCLUSIONS

1. Coatings and powder metallurgy composites of metal-bonded chromium carbide with dispersed silver and  $\text{CaF}_2/\text{BaF}_2$  eutectic are self-lubricating materials from low temperatures to 900° C. The best composition so far evaluated is by weight percent: 70 bonded carbide-15 silver-15 fluoride

eutectic. The coatings are designated as PS212 and the powder metallurgy compacts as PM212.

2. The coatings and the free-standing powder metallurgy compacts are complementary in their applicability. The powder metallurgy processes can easily produce parts with surfaces that are very difficult to spray as, for example, small bore cylindrical bearings.
3. The compressive strength of sintered PM212 from room temperature to 760° C compares favorably to the room temperature strength of many common grades of bronze and carbon sliding contact bearing materials. HIPed PM212 is about three times stronger than the sintered version, probably because it is fully dense while the sintered material contains 20 percent porosity.
4. The friction coefficients of PS212 decrease with increasing sliding velocity making them promising candidates as high-speed, high-temperature sliding contact seal materials.

#### REFERENCES

1. Sliney, H. E.: U.S. Patent No. 4,728,448 Carbide/Fluoride/Silver Self-Lubricating Composite, Mar. 1, 1988.
2. Sliney, H. E.: The Use of Silver in Self-Lubricating Coatings for High Temperatures. ASLE Trans., vol. 29, no. 3, July 1986, pp. 370-376.
3. DellaCorte, C.; and Sliney, H. E.: Composition Optimization of Self-Lubricating Chromium-Carbide-Based Composite Coatings for Use to 760° C. ASLE Trans., vol. 30, no. 1, Jan. 1987, pp.77-83.
4. DellaCorte, C.: Tribological Composition Optimization of Chromium Carbide-Based Solid Lubricant Coatings for Foil Gas Bearings at Temperatures to 650° C. Surf. Coat. Technol., vol. 36, no. 1-2, 1988, pp. 87-97.
5. DellaCorte, C; and Sliney, H. E.: Tribological Properties of PM212: A High Temperature, Self-Lubricating Powder Metallurgy Composite. NASA TM-102355, 1990.
6. Sliney, H. E.: Hot Piston Ring/Cylinder Liner Materials - Selection and Evaluation. SAE Paper 880544, also DOE/NASA/50122-73, NASA TM-100276, 1988.

TABLE I. - DATA SUMMARY

Temperature, °C	Load, kg	Friction coefficient, $\mu$ m	KPM212, $\text{mm}^3/\text{Nm}$	KRené 41, $\text{mm}^3/\text{Nm}$	Number of tests
850	0.5	$0.29 \pm 0.03$	$4.1 \pm 2.0 \times 10^{-6}$	$5.0 \pm 1.0 \times 10^{-6}$	2
760	↓	$.35 \pm 0.06$	$3.6 \pm 0.9 \times 10^{-6}$	$1.0 \pm 6 \times 10^{-5}$	4
<sup>a</sup> 350		$.38 \pm 0.02$	$3.9 \pm 1.8 \times 10^{-5}$	$3.5 \pm 1.0 \times 10^{-6}$	3
25		$.35 \pm 0.05$	$3.2 \pm 1.5 \times 10^{-5}$	$7.0 \pm 2.0 \times 10^{-5}$	2

<sup>a</sup>Friction and wear increase after  $\approx 8$  km of sliding at 350 °C. Friction increases to  $\approx 0.5$  and the pin material transfers to surface of disk.

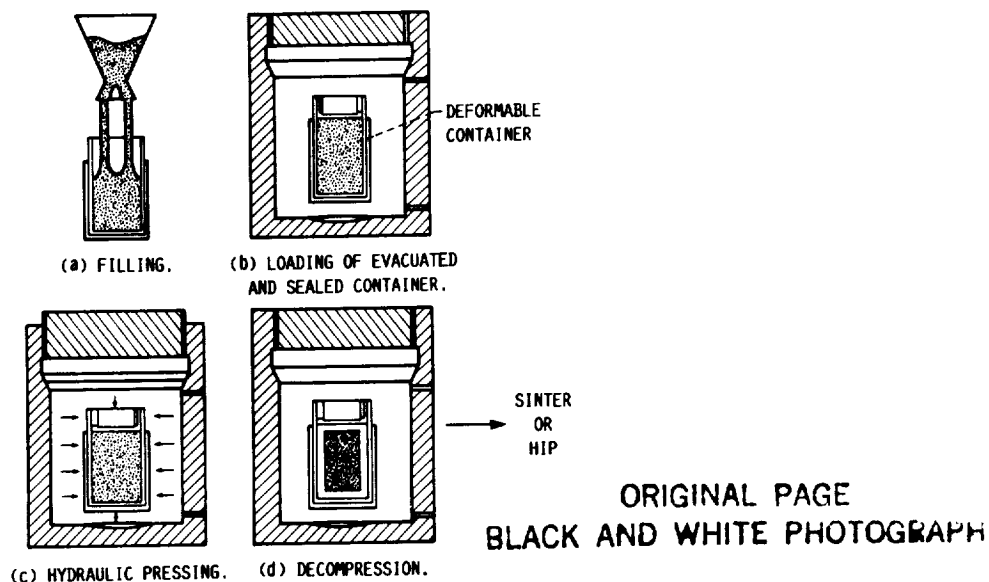


FIGURE 1. - STAGES IN COLD ISOSTATIC PRESSING.

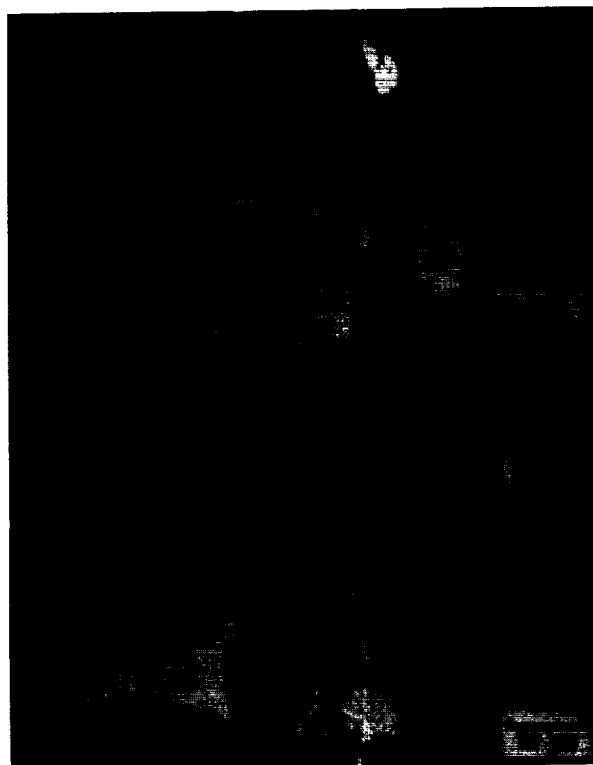
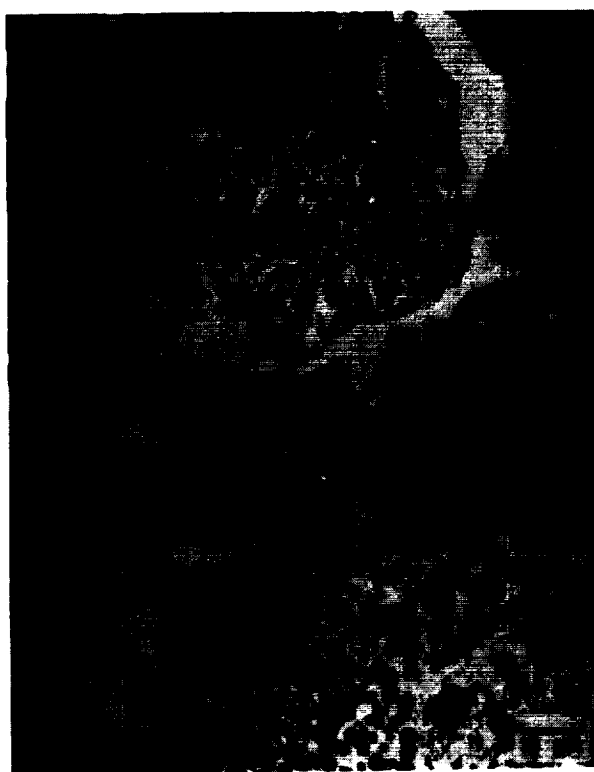


FIGURE 2. - DISTRIBUTION OF COMPONENTS IN SINTERED PM212.

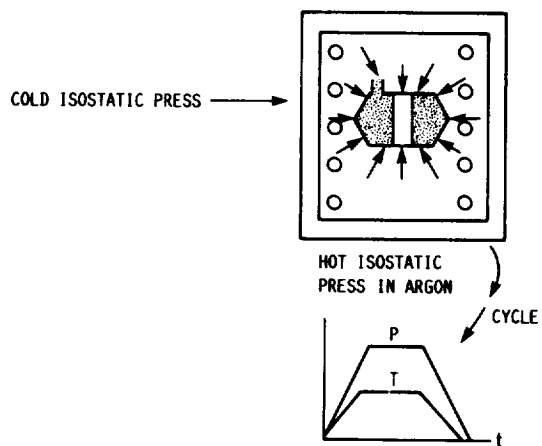
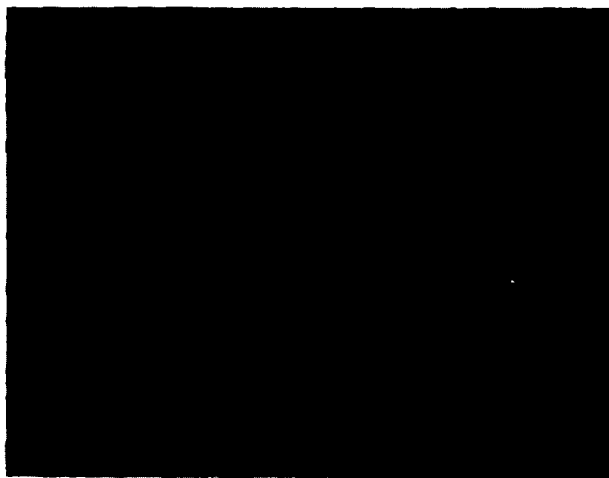


FIGURE 3. - STAGES IN HOT ISOSTATIC PRESSING.

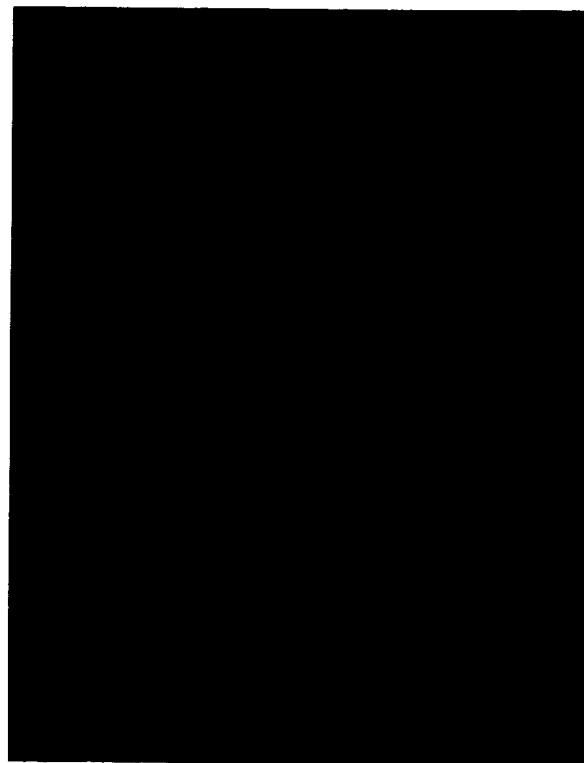


(a) PM212-CS.



(b) PM212-CH.

FIGURE 4. - ILLUSTRATION OF DENSITY DIFFERENCE.



(a) VERTICAL ILLUMINATION.



(b) OBLIQUE ILLUMINATION.

FIGURE 5. - HIPPED PM212 MICROSTRUCTURE BY OPTICAL MICROSCOPY.

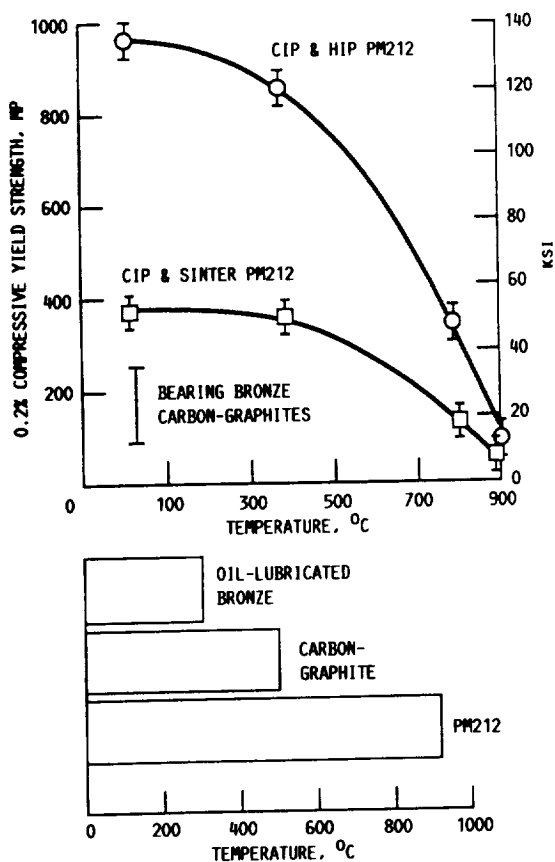


FIGURE 6. - STRENGTH AND MAXIMUM SERVICE COMPARISON OF PM212 WITH CONVENTIONAL BEARING MATERIALS.

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



FIGURE 7. - PM212 PARTS FABRICATED BY EDM.

ORIGINAL PAGE IS  
OF POOR QUALITY

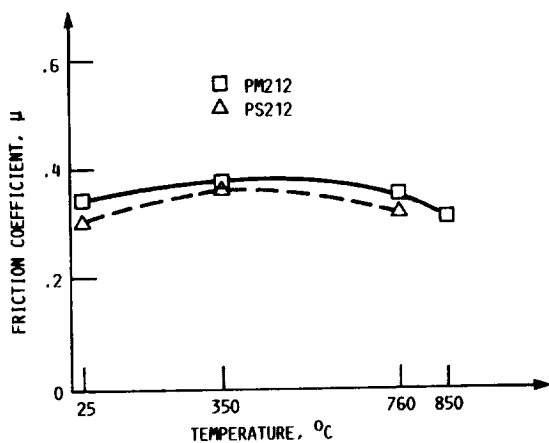


FIGURE 8. - FRICTION COEFFICIENT COMPARISON OF PM212 AND PS212 (PLASMA SPRAYED COATING) IN AIR, 35 PERCENT R.H. AT 25 °C, 0.5 kg LOAD, 2.7 m/s SLIDING VELOCITY.

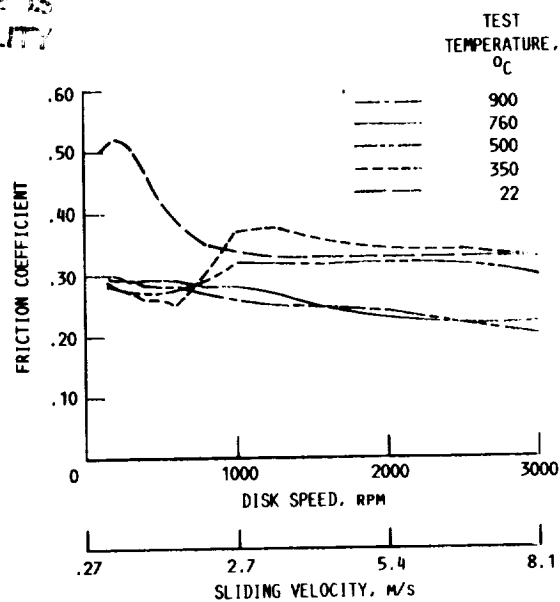


FIGURE 9. - FRICTION COEFFICIENT VERSUS SLIDING VELOCITY, 0.5 kg LOAD, IN AIR, 35 PERCENT R.H.

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH

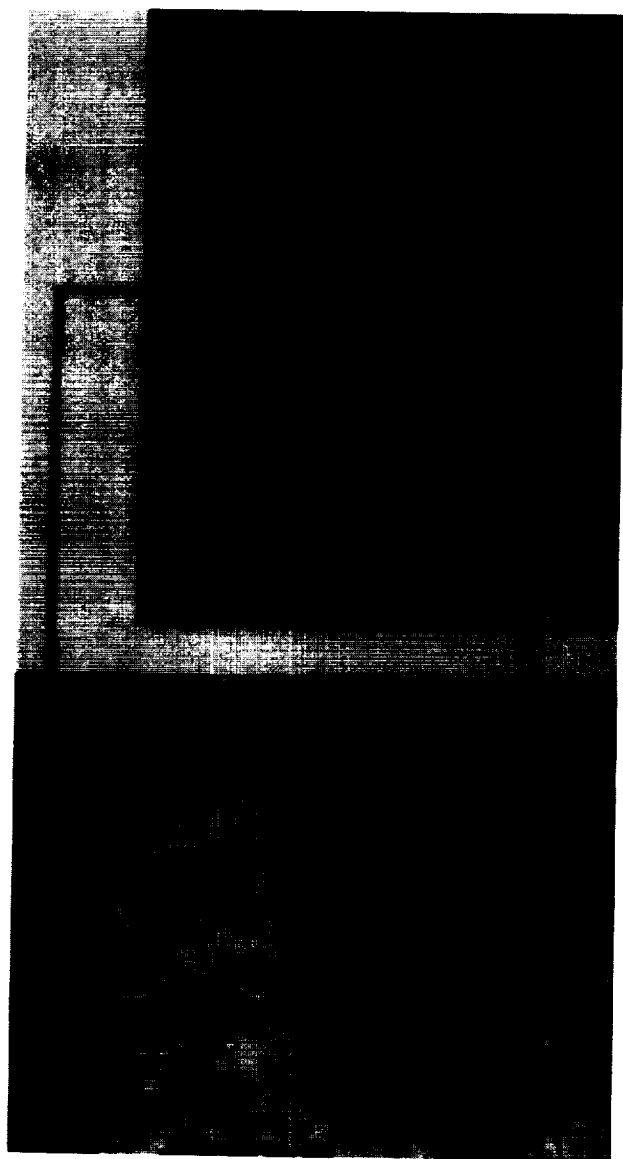


FIGURE 10. - SURFACE DENSIFICATION OF PM212-CS ON A WEAR SCAR.

ORIGINAL PAGE IS  
OF POOR QUALITY

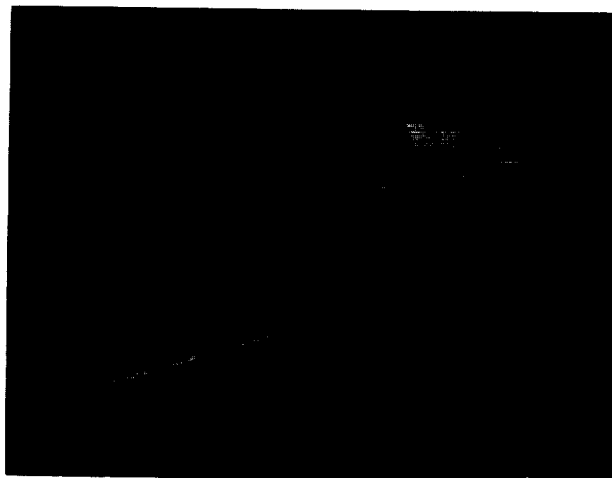


FIGURE 11. - HIGH SPEED GAS BEARING JOURNAL AND THRUST PLATE COATED WITH PS212.



FIGURE 12. - AUTOMOBILE STIRLING ENGINE.

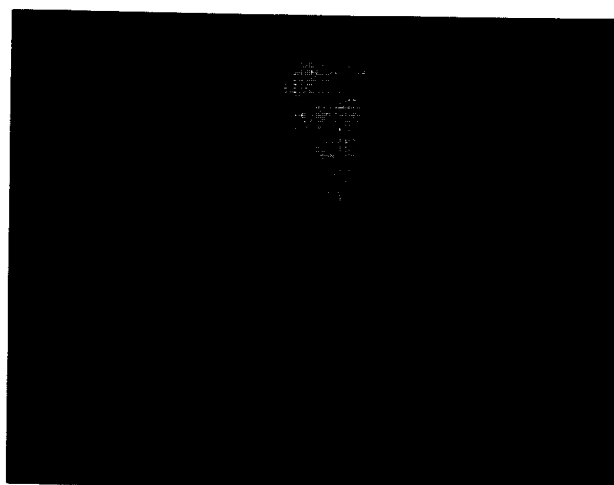


FIGURE 13. - HIGH SPEED SHAFT SEAL WITH PS212 COATING ON SEAL SURFACE.



## **SESSION E - MANUFACTURING AND FABRICATION TECHNOLOGY**

**Tuesday November 27, 1990**

- **Robotics In Space-Age Manufacturing**
- **Variable Polarity Plasma Arc Welding**
- **High-Pressure Water Jet Cutting And Stripping**
- **Cost-Efficient Manufacturing Of Composite Structures**
- **Rapid Induction Bonding Of Composites, Plastics, And Metals**
- **Reliability And Risk Assessment Of Structures**
- **A Semi-Automated Process For The Production Of Custom-Made Shoes**
- **Lightweight, Fire-Retardant, Crashworthy Aircraft Seat Cushioning**



**ROBOTICS IN SPACE-AGE MANUFACTURING**

Chip Jones  
Process Engineering Division  
Materials & Processes Laboratory  
Marshall Space Flight Center, Alabama 35812

**ABSTRACT**

The Marshall Center is developing robotics technologies to improve manufacturing of space hardware. This paper will cover application of robotics to:

- Welding for the space shuttle and space station Freedom programs
- Manipulation of high-pressure water for shuttle solid rocket booster refurbishment
- Automating the application of insulation materials
- Precision application of sealants
- Automation of inspection procedures

Commercial robots are used for these development programs, but they are teamed with advanced sensors, process controls, and computer simulation to form highly productive manufacturing systems. Many of the technologies are also being actively pursued in private sector manufacturing operations.

**INTRODUCTION:**

Millions have watched the astronauts use the space shuttle's Remote Manipulator System (RMS), but NASA's use of robotics is not limited to space. Here on earth, industrial robots are being used to manufacture and refurbish components for the space shuttle. In these applications, the robots have been found to reduce costs and to improve the quality of the conventional processes that they replaced. New applications, presently under development, promise to extend the advantages of robotic process automation to other shuttle components, as well as the space station Freedom.

In this discussion, some of the applications of commercial robots in manufacturing for the space program will be presented. It is hoped that some of the technology and approaches that are used by NASA will advance related private-sector industries.

**BACKGROUND:**

Even as the first flight of the space shuttle was underway, engineers at NASA's Marshall Space Flight Center were beginning a program that would radically change the approach to building launch vehicle hardware. Since the new Shuttle's goal was to allow more frequent space flight at a lower cost, the old way of fabricating the flight hardware by largely manual methods was no longer practical. For this reason, the Productivity Enhancement Complex at MSFC was created to develop improved methods of manufacturing to reduce costs and improve consistency. This was done by bringing computer automation to bear on as many manual manufacturing applications as possible. To that end, modern industrial robots would come to play a vital role.

Most people would imagine that robots used for manufacturing tasks would be best applied in situations where a large number of parts have to be produced. Pictures of spot-welding robots in an auto plant producing thousands of car bodies and high-speed assembly robots fabricating electronic circuit boards probably come to mind. With the ability to lift loads exceeding 50 kilos and place them to within one millimeter of a programmed location without tiring, industrial robots have earned a prominent place in the manufacture of high-volume goods.

Other features of robots make them attractive for use in space flight hardware manufacturing. The ability to work in environments hazardous to humans makes them ideal for spraying insulating and

coating materials, as well as the removal of these same materials. The control computers of modern industrial robots can coordinate the motion of as many as a dozen axes of motion while controlling a complex manufacturing process such as arc welding. This allows very consistent results even over very long processing times, improving overall quality by reducing the effects of operator fatigue and inattention. The universal programmability of a robot arm can allow the use of one piece of equipment on many applications and parts. In the past, the common way to mechanize operations was to design a special machine for each application, resulting in prohibitive equipment cost for small operations. The speed and endurance of a robot arm, coupled with computer data communications has made possible continuous inspection schemes, giving engineers more reliable data from which to determine the flight readiness of a part.

Before examining the areas of robotic activity it may be useful to briefly review the Shuttle Vehicle and the Space Station Freedom configurations. Unlike previous NASA vehicles such as the Mercury, Gemini, and Apollo series that used totally expendable launch systems, the Shuttle and most of its elements are completely reusable. Thrust for lift-off is provided by the two solid rocket boosters (SRB's) and the three Main Engines aboard the Orbiter. Fuel, in the form of liquid oxygen and hydrogen for the Main Engines is supplied by the large External Tank (ET). After the SRB's expend their fuel, they are jettisoned and recovered for re-use on a future launch. When the orbiter has reached the appropriate altitude and velocity, the Main Engines are turned off and the External Tank released. The ET is the only element of the Shuttle's launch cluster that is not recovered at the present time, being destroyed by re-entry into the earth's atmosphere. A major feature of the Shuttle's unique capabilities is its re-usability and land-based recovery.

#### **PRESENT ROBOT APPLICATIONS:**

Industrial robots are now used in flight hardware manufacturing for the application and removal of thermal protection systems for the Shuttle Solid Rocket Boosters, and for robotic welding of the Shuttle's Main Engines.

#### **THERMAL PROTECTION PROCESSING**

Thermal protection systems (TPS) are an integral part of any space vehicle and the Shuttle is no exception. The Orbiter itself is equipped with a covering of insulating tiles to protect it from the heat of re-entry. The SRB's and the ET must also be protected. The ET, though expendable, must be insulated to reduce boil-off of its cryogenic contents and to reduce the accumulation of ice prior to launch. The skin of the SRB's must be protected from aerodynamic heating during the ascent phase of flight and from the shock of hitting the water after parachuting back to earth. After recovery of the SRB's all of the TPS material must be removed, then re-applied before their next use.

The thermal protection system used for most of the SRB's is called the Marshall Sprayable Ablator (MSA). MSA is an epoxy-terminated urethane resin filled with various organic fibrous ingredients for strength. The result is a low-density ablator with the necessary insulation properties. It is applied to the SRB components by spraying, which replaces hand-layup of sheet cork material. This is accomplished at the Kennedy Space Center with a robot working in concert with a rotary table on which the part being sprayed is located. The robot and turntable are housed in a spray cell that coordinates the activity of the robot as well as the precise mixing of the components of the insulating material. The robot cell has been in operation since 1980 after favorable results were obtained from an MSFC prototype system.

Implementation of the four cells at KSC was by United Space Boosters, Inc. Three cells utilize six-axis gantry robots, with rotating tables. The motion of the tables are coordinated with the robot motion, allowing a component much larger than the robot to be sprayed using one setup (see figure 1).

Two TPS application cells perform a variety of tasks on the SRB nose cone, forward skirt, frustum, and aft skirt, as follows:

- Pre-cleaning of parts, using hot water spray
- Spraying of the MSA material, using computer-controlled mixing equipment.
- Sanding of cured MSA, to smooth the surface.
- Application of a white hypalon sealant, for environmental protection.

The major incentives for using the robots for these processes are:

- Protecting humans from the toxic propellants used for spraying.
- Precise process control, made possible by having all mixing and spraying operations under computer control.
- Ability to reach all parts of the component during a spraying operation.

#### **TPS REMOVAL**

Complementing the robotic facilities for MSA application are the MSA removal cells. These systems remove the MSA material from the SRB's after their post-launch recovery. Removal of the partially-ablated material is necessary before re-application for the next flight.

A narrow, high-pressure stream of water is directed at the surface of the SRB component to strip away the old material. In the past, a rubber-suited worker would manually sweep a hand-held "car wash" type wand across the surface until sufficient material was removed. Due to the high pressures and the attendant reaction forces, worker fatigue was a problem. In addition, because a human could not accurately maintain a constant distance and surface feedrate, many spots would be missed, requiring another touch-up pass.

The introduction of a robot to perform this task allowed the stripping process to be controlled to a degree unattainable by its human counterparts. The consistent motion of the robot arm and integration with the control of the pumping equipment meant fewer touch-ups and less chance of damage to the underlying part. In operation, the system can be programmed to remove only the MSA, then pass over again to remove successive layers of paint and primer. Cleaning to the bare aluminum substrate is possible without damage.

The refurbishment facility at the Kennedy Space Center utilizes a NIKO gantry-type robot in concert with a part turntable to strip individual components of the SRB's. A new robot system, built at MSFC for stripping of the boosters before disassembly, is composed of a GMF robot mounted on a motorized transporter, to give the robot access to all surfaces. The systems operate according to pre-programmed motion and processing parameters.

#### **ROBOTIC WELDING**

Eight robots are presently in use welding space shuttle main engines (SSME) at the Rocketdyne facility in Canoga Park, California. Six Cybotech H-8 robots weld small-to-medium-sized parts such as pump housings, and two Cybotech H-480 robots weld larger parts, including the engine's exit nozzle, which is slightly over three meters long and 2.6 meters in diameter. All robot stations have seven axes of motion and are equipped with two-axis part positioning tables.

SSME design calls for numerous welded joints, with each engine requiring approximately 530 linear meters of welds, on materials such as nickel-based alloys, stainless steel, and titanium alloys. Thicknesses range from 1.5 to 12 millimeters. The Gas Tungsten Arc (GTA) welding process is used by the robots. Some 75% of robotic welds are required to meet the most stringent requirements for SSME post-weld inspection, where X-ray, dye-penetrant, and ultrasonic methods must be used.

Thicknesses range from 1.5 to 12 millimeters. The Gas Tungsten Arc (GTA) welding process is used by the robots. Some 75% of robotic welds are required to meet the most stringent requirements for SSME post-weld inspection, where X-ray, dye-penetrant, and ultrasonic methods must be used.

While it is undeniable that automated welding produces more consistent high-quality results, most welding on the SSME was done by hand. Welds with simple geometries, such as linear or circular paths in a single plane had used mechanized equipment, with good results, but this kind of application was limited. Since each engine had more than 4000 distinct welds, with many requiring complex torch motion, the cost to conventionally automate these was wildly prohibitive. In 1981, NASA engineers started investigating robots as a potential way around this problem.

A program was initiated develop robot controls to make them practical for the welding required on the SSME. The major development tasks were as follows:

- Specialized welding process control computers had to be integrated into the robot's motion control, since the welding commands available from conventional robots could not control all the parameters required for aerospace-quality welds.

- New methods for designing tooling and fixtures were developed so that the position of the weld joint was known by the robot.

- Improved robot programming techniques were needed to simplify the task for welding engineers and technicians.

Robotic welding on the SSME was implemented in production in 1986 after a coordinated development between NASA and Rocketdyne at Marshall Space Flight Center and at the production facility at Canoga Park. As benefits, the resulting welds:

- Require less joint preparation.

- Exhibit reduced distortion

- Produce fewer process-related defects, such as crater-cracking, even on crack-prone alloys.

Much of the credit for these benefits can be attributed to simply having the welding process under computer control, where the optimum parameters for a weld can be consistently repeated. The robot can control the weld at higher current levels, penetrating through thick metals without requiring a beveled edge. Reduced distortion on welded parts has resulted, since fewer weld passes are required for a given material thickness.

Another critical element has been advanced coordinated motion control, that maintains a constant torch velocity along the joint. The coordinated motion between the torch and part positioner has allowed the more complex parts to be welded in a preferred orientation. Even on very convoluted parts, the robot and positioner move together to follow the joint without resorting to "out of position" welding, which requires adjustments to the welding parameters to avoid molten metal dripping from under the torch.

More than 300 weld types have been converted to robotics to-date. These welds, though a small percentage of the total number per engine, are a high percentage of the more difficult and lengthy welds. Nearly fifty percent of the total length of GTA welds on the SSME are now being made robotically. As a result, weld quality acceptance has improved significantly.

## **DEVELOPMENT ACTIVITIES**

Other areas of robotic activity are presently under development that promise to expand the use of robots in the production of NASA's space flight hardware.

### **SPACE STATION ROBOTIC WELDING**

Rather than replace a conventional operation with a robotic one, the space station Freedom project is considering the use of robots for manufacturing from its inception. Welding of space station structural elements will be made using a robot utilizing two part positioners and the Variable Polarity Plasma Arc welding process. Approximately one-third of welds for the common module, logistics module, docking node, and cupola will be made with this system (see figure 2). Thirteen joints representing seven different configurations will be accommodated by the one system, avoiding the need for some five conventional dedicated tools. The robot also allows all welds to be made in the same orientation to gravity by use of coordinated motion. This feature will greatly simplify weld process development. Production robotic welding is scheduled to start in the fall of 1991.

### **WELDING SENSORS**

Sensors, incorporated into the operation of the robots, are being developed for SSME robotic welding. The large number of critical welds and the high part value of the engine components makes this an ideal application for feedback-controlled robotic welding. Rocketdyne will implement a pair of sensors, one to guide the torch accurately over the weld joint, the second to ensure correct weld penetration into the metal.

The seam-tracking sensor guides robot motion to follow joints to within .25mm of the seam. It adjusts torch position to correct for:

- Inaccuracies in the robot program
- Variation in part fit-up
- Thermal distortion of the part during welding

The basic element of the seam tracker is a through-the-torch coaxial-view welding torch. A miniature closed-circuit tv camera in the top of the torch peers through the torch body to relay an image of the joint to the image processing computer. An electro-optic light valve in the torch adjusts the image intensity reaching the camera to compensate for varying light levels as welding current changes. Machine vision software in the computer analyzes the image, locates high-contrast points and determines the location of the seam. Corrections to the robot's path are then sent to the robot controller. The result is an automatic seam tracker and visual weld monitor that does not add any size to the weld torch, important for welding in limited access areas typical to the engine.

The penetration control sensor measures the natural oscillation frequency of the molten weld pool by monitoring light intensity from the weld pool area with a photosensor. This frequency corresponds to weld penetration, characterized by weld back-side bead width. Current pulses cause the pool to oscillate; if the bottom of the pool is solid - not penetrated through the thickness of the part - the pool is stiff and oscillates rapidly at low amplitude. As the pool grows and penetrates with increasing power to the torch, frequency decreases and amplitude increases. A computer analyzes the oscillation frequency and amplitude then sends adjustments in the weld current to the process controller.

The welding sensors are in the final stages of testing on the production floor. Rigorous qualification procedures will accompany implementation. Production welding by sensor-guided robots should begin in 1991.

## MEASUREMENT SYSTEMS

The speed and accuracy of a robot's motion make it ideal for automating measurement and inspection tasks.

In one application investigated at MSFC, a robot was used to scan a contamination monitor over the inside of the solid rocket motor case to check for minute levels of contamination from grease and similar manufacturing residue. This residue could inhibit bonding of the fuel to the inside of the motor. The sensor measured the optical emission of contaminants stimulated by its ultraviolet light source. The light source is directed at the surface as the robot scans the component. A map of contamination levels was generated by correlating the location of the robot to the feedback from the sensor, giving engineers more complete information on which to base a decision to re-clean part of the surface. Manual scanning of the part was extremely time consuming, since the part is too large (four meters diameter by three meters high) to reach in one setup. The accuracy of the manual process is some fifty times less sensitive than the robotic method. The sensor that the robot uses requires precise control of the standoff distance, which cannot be maintained manually without touching the part. A written record of the part's cleanliness level is an added benefit for quality history. This sensing method has been implemented for cylindrical components using a lead-screw type manipulator. A robotic system for more complex geometries is being assembled at the production facility.

Another measurement application under development is to map the thickness distribution over the external tank's aluminum skin. At the pointed ends of the tank, the skin must be chemically milled to the proper thickness. This is now a slow, manual process, where the panel's thickness is measured in selected places, and a contour map drawn by hand on the skin in order to specify the required extent of masking on the surface. Once the mask over the excessively thick areas is removed, the chemicals can reduce these areas to their specified thicknesses. Use of a robot allows the whole panel to be mapped in one operation using an ultrasonic thickness gauge. An offline computer correlates the robot's position to the local thickness to produce a program to tell the robot where to scribe contour lines for removing the maskant.

## SIMULATION AND OFFLINE PROGRAMMING

Very exciting areas of development in robotics are simulation of a robot's motion and offline programming.

Space flight hardware, in many instances, can have more value than the robot performing the task. For example, when some of the robotic welds are made on the main engines, more than one million dollars worth of investment has already been incorporated in each part. For this reason, everyone must be assured that the robot will perform its task right the first time. Unfortunately, a trial part to practice on is very hard to obtain. In these cases, a CAD/CAM system is used to simulate the operation long before any money is invested in robot hardware. Before any large robots were delivered for welding the large SSME components, a computer simulation was conducted, making typical welds on the components for which the system was intended. This exercise found the ideal placement for the part vis-a-vis the robot, assuring that the robot could reach all welds and avoid collisions. Part tooling and facilities placement were well under way before the robots were delivered.

In the case of welding for the Space Station radial docking port structure, simulation and offline programming were essential for successful implementation. Due to the part's large size, workcell layout is critical. A 10 centimeter change in the location of the robot, part, or positioner could render the weld seam unreachable or cause a collision. The complex shape of the path requires precise programming of the path to assure that the seam is closely followed and the resultant simultaneous motion of the part and weld torch is smooth. Use of the design model of the part in calculating the motion reduces the error. Manually programming the path, based on our experience, would take more than a week, and would be very hard to modify if a slight change needed to be made.



Computer simulation proved indispensable for implementation of the mobile robot used in refurbishing solid rocket booster parts. Because of budgetary constraints, a robot large enough to reach every part on the booster was not feasible. A six-axis robot with an average-sized envelope was used. Computer kinematic simulations were used extensively to determine the proper position of the robot transporter to be sure that the areas that needed to be sprayed could be reached. Since an actual booster was not be available for teaching the robot programs manually, offline programming software was used to calculate every motion needed. Computer-generated pictures of the operation were created from the operator's viewpoint so that the robot position could be verified prior to program execution. All these programs were created at MSFC and transmitted to the Kennedy Center facility.

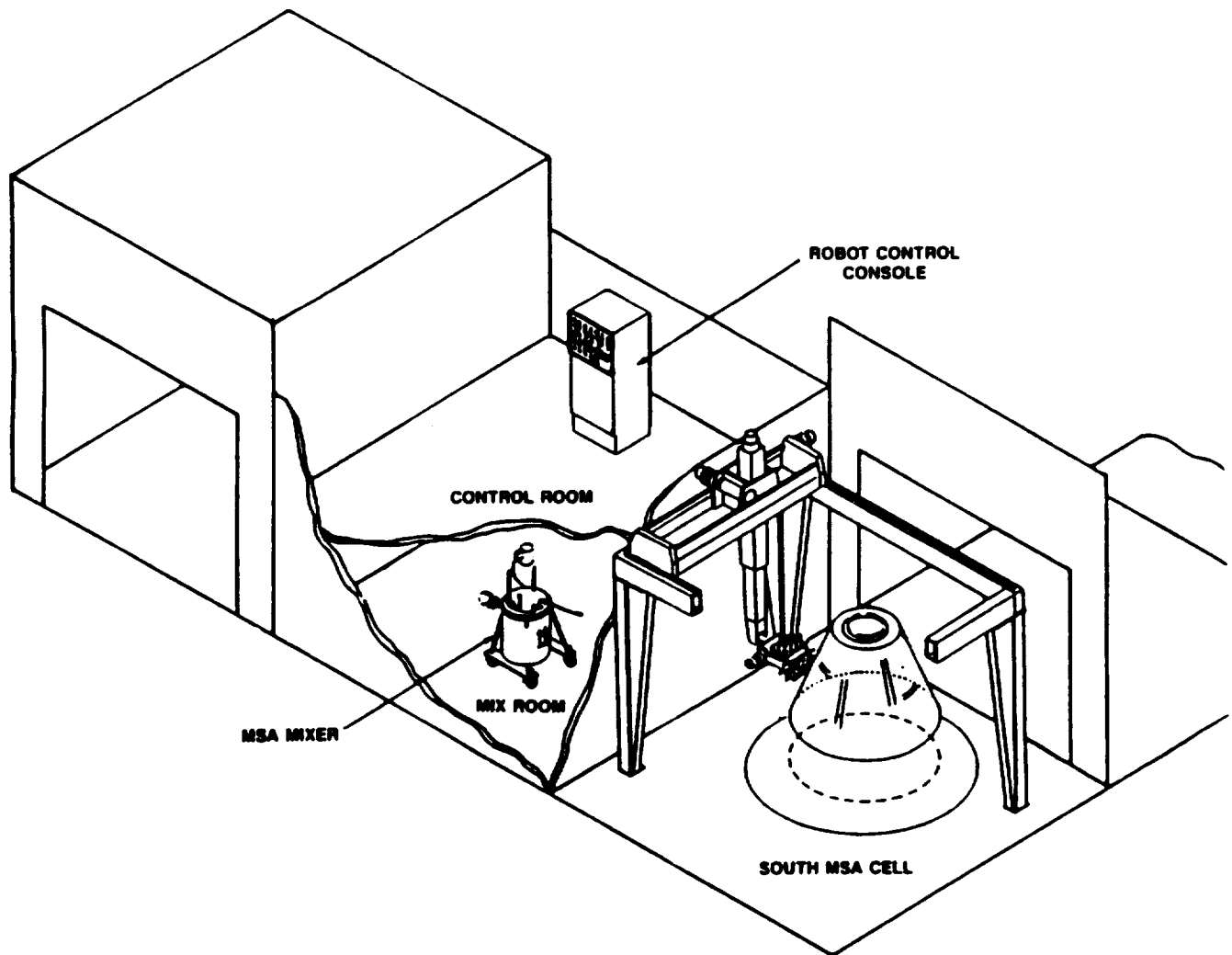
Thus, robot kinematic simulation and offline programming has proven to be especially critical to aerospace applications where many programs need to be written for just a few parts, which may not be readily available. It can also be valuable to determine the range of tasks that can be applied to limited robot resources. Control of the programs used by the robot in production can be improved, for better quality control.

### **SUMMARY**

The National Aeronautics and Space Administration has identified robotic technology for automation of manufacturing tasks for the space shuttle and space station Freedom. Although these tasks differ from traditional applications of industrial robots, they have proven to reduce costs while improving quality.

### **REFERENCES**

1. K. Fernandez, et. al., "NASA's Use of Robotics in Materials Processing", IEEE -IAS Annual meeting, 1982.
2. D. Gutow, M. Smith, "Sensors Guide Shuttle Welding," Welding Design and Fabrication, May, 1990.
3. D. M. Shelley, "Robotic CPI Team Activity," Rocketdyne Internal Letter, Sept. 1990.
4. C. Jones, D. Stiles, "Workcell Calibration for Effective Offline Programming," Conference Proceedings, SME, Robotics in Aerospace Manufacturing, February, 1989.



**Figure 1:**  
**TPS Application System**

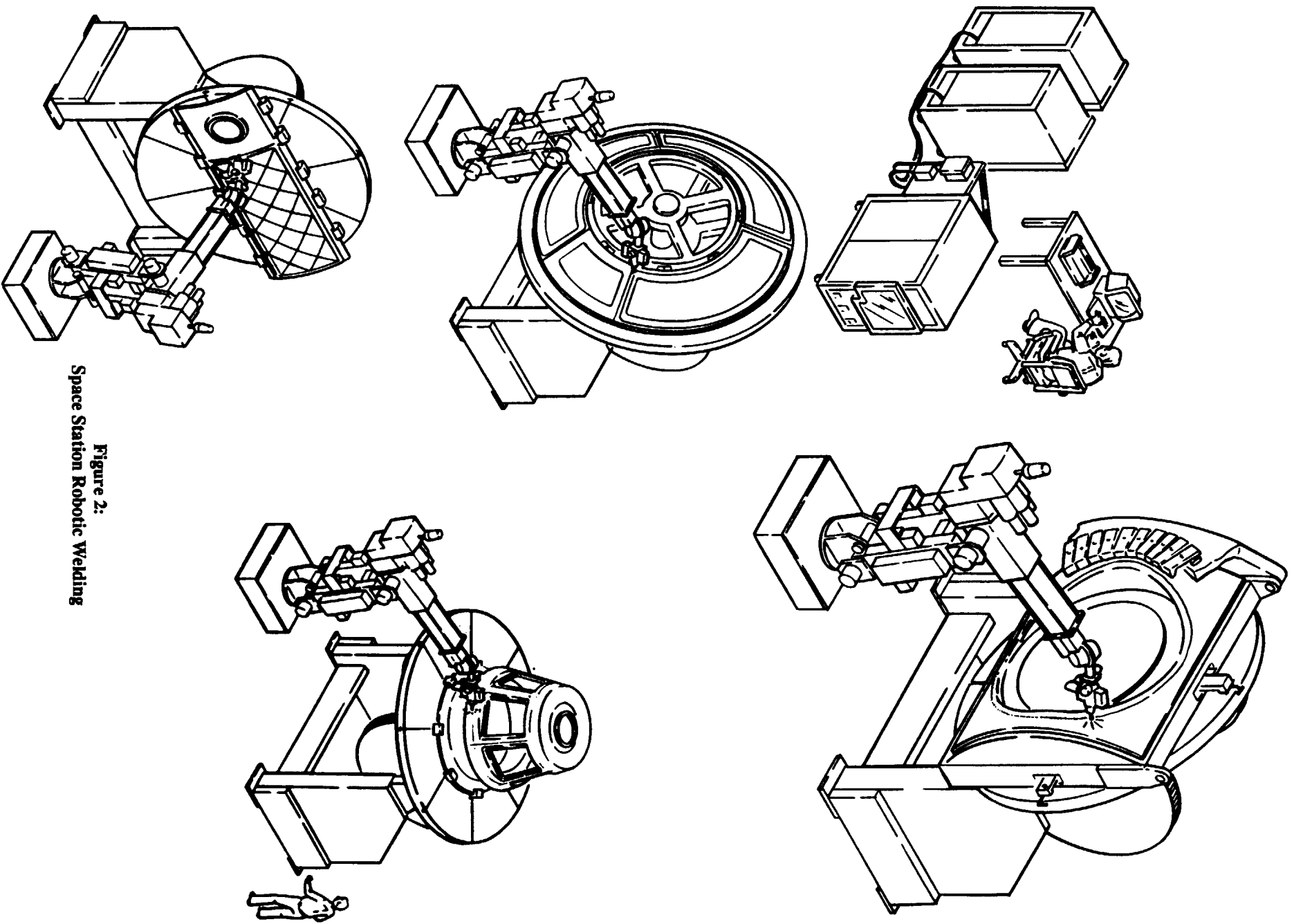


Figure 2:  
Space Station Robotic Welding

## VARIABLE POLARITY PLASMA ARC WELDING

E.O. Bayless, Jr.  
Marshall Space Flight Center

Technological advances generate within themselves dissatisfactions that lead to further advances in a process that goes on and on to the benefit of all concerned. Today I intend to describe very briefly a series of advances in welding technology which culminated in the Variable Polarity Plasma Arc (VPPA) Welding Process and an advance instituted to overcome the latest dissatisfactions with the process: automated VPPA welding.

### THE TROUBLE WITH GTA WELDING

Gas-shielded Tungsten Arc (GTA) was developed during World War II for welding light aircraft alloys of aluminum or magnesium. The GTA process arc is embedded in an inert gas that flows around a nonconsumable tungsten electrode as shown in Figure 1a. The inert gas protects the surface of the hot metal from reacting chemically with the air. This process was used to fabricate the Saturn V, and, until recently, the Space Shuttle External Tank.

One source of dissatisfaction with GTA welding is the excessive width of deep penetration welds. The force of the GTA arc isn't enough to penetrate the molten weld metal. The heat enters the weld from the arc at the surface of the puddle, and the heat spreads out from the heat source in a roughly spherical pattern. Although fluid circulation in the weld puddle can vary this, the width of a deep penetration weld bead is about the same on each side as the depth of penetration. Thus heavy welds are made with multiple beads. Each bead presents new opportunities for defects to enter the weld.

Another source of dissatisfaction with GTA welding is its sensitivity to contamination. A little grease or water can generate hydrogen, which aluminum avidly absorbs and then regurgitates as porosity upon solidification.

Reverse polarity welding, where the tungsten electrode is made positive with respect to the workpiece so that a cleaning action takes place as positive ions strike the workpiece, reduces sensitivity to contamination, but at a price.

A bit of explanation is due here. A welding arc consists of a plasma of ionized gas. There are neutral inert gas atoms, positive inert gas ions, and negatively charged electrons. The positive and negative charges balance, so the plasma is neutral. The positively charged ions drift towards the negatively charged cathode. From the sporadic flashing seen at the cathode it would appear that the positive ions charge up the surface oxide to the point of electrical breakdown and that the observed cleaning action may take place by a succession of electrical breakdown explosions on the cathode surface.

The electrons in the arc plasma drift towards the positively charged anode. They drift much faster than the heavy positive ions and carry the bulk of the arc current. They generate appreciably more heat where they enter the anode than where they leave the cathode. Consequently GTA operations are normally carried out such that the workpiece is the anode and gets the bulk of the heat. This is "straight polarity." In "reverse polarity", where the workpiece is made the cathode, the electrode receives the bulk of the heat. Therefore the penetration in reverse polarity has to be greatly curtailed to keep the electrode from burning up. Welds on the Saturn V or the Space Shuttle External Tank were generally made using straight polarity (for the sake of the greater penetration) compensated for by very stringent cleaning procedures (scraping, white gloves, etc.). The unavoidable defects were ground out and repaired.

## **THE TROUBLE WITH PLASMA ARC WELDING**

Plasma Arc Welding (PAW) was developed in the 60's. An early application was the fabrication of steel (D6AC) rocket cases by Westinghouse. In the PAW process the arc is emitted through a water cooled nozzle surrounding the tungsten electrode as shown in Figure 1b. The hot gas of the plasma expands through the nozzle in a high speed jet that easily penetrates the molten metal in the weld puddle so as to deliver heat throughout the thickness of the workpiece and not just at its surface. Deep, narrow single pass welds become feasible. In fact, in PAW practice welding in the "keyhole" mode with complete penetration of the workpiece is usual. With a high speed plasma jet continually flushing out the weld keyhole the sensitivity of the PAW process to surface cleanliness is greatly reduced.

But the PAW process is unsatisfactory when it comes to dealing with tenacious oxides. In straight polarity aluminum alloys cannot be welded. One encounters an invisible surface film that prevents the liquid metal from flowing together into a sound weld. One does not get a weld bead; one gets a crinkly mess where the metal has been melted. If polarity is reversed to take advantage of the effect of the cleaning effect of reverse polarity, this problem disappears, but penetration capability is drastically reduced.

## **THE TROUBLE WITH VARIABLE POLARITY PLASMA ARCH WELDING**

VPPA welding was developed in the late 60's and early 70's by B.P. VanCleave at Boeing Company. This process resembles the PAW process, except that the VPPA process incorporates a variable current waveform. The system spends part of its time in reverse polarity for the cleaning benefit and the rest of its time in the more efficient straight polarity mode. For 2219 aluminum about 4 milliseconds reverse and 19 milliseconds straight polarity has been found to work well. A special power supply was developed for this process on contract by Hobart Brothers Company as existent power supplies were not up to the demands of the process for the heavier welds.

Marshall Space Flight Center (MSFC) began to develop a VPPA system in the late 70's, shortly after the emergence of Hobart's VPPA power supply, with a view towards replacing the GTA system then in use for fabrication of the Space Shuttle External Tank. In 1984 an article was published in the *Welding Journal* (September 1984, pp.27-35) describing the development of the VPPA system at MSFC and noting that, "Implementation of the VPPA welding technique in the production of the Space Shuttle External Tank is nearing completion." A picture of the present VPPA system at MSFC is shown in Figure 2. At present the implementation is 80% complete.

To operate the GTA system one has to set current, standoff or voltage, weld speed, shield gas flow, and wire feed rate: 5 variables. To operate the VPPA system one has to set addition reverse polarity current, reverse and straight polarity time cycle increments, plasma gas flow and plasma jet orientation: 5 more variables making a total of 10 variables. The initial system as described in 1984 was computer controlled. General Digital Industries, Inc. of Huntsville, Alabama, set up the digital control system as subcontractor to Hobart. The heart of the system was the PAL-100 computer (PAL = "Process Automation Language") operating with a Digital Equipment Corp. (DEC) LSI-11/23 microprocessor. With this equipment the welding variables can be programmed in advance. The operator will still have to make trim adjustments, however, to compensate for local perturbations. The operator will also have to make seam tracking corrections and wire entry angle adjustments as needed. This can result in defects or, at least, variability of weld properties. Considering the very high cost of the parts being processed and the repair costs, this is unsatisfactory.

## **THE AUTOMATED VARIABLE POLARITY PLASMA ARC WELDING SYSTEM**

A brief description of current efforts to automate the VPPA welding system completely and to eliminate the hand of the welder on the controls entirely has been given in the Research and Technology 1989 Annual Report of the Marshall Space Flight Center (C. Kurgan: Fully Automated Variable Polarity Plasma Arc Welding, p. 240). A schematic view of the effort is shown in Figure 3. The work comprises development of sensors to feed back information and control in real time with respect to the weld seam (tracking) and the

configuration of the weld bead/puddle/keyhole, development of a mathematical model of the relation of the system parameters to the bead configuration, and special (e.g., wirefeed and torch rotation) controls as well as general system control algorithms. We envision this system as the future standard VPPA system for aerospace fabrication. It is expected to reduce defects to the extent that radiographic inspection requirements can be relaxed.

#### **UTILIZATION OF THE VPPA SYSTEM: PRESENT AND FUTURE**

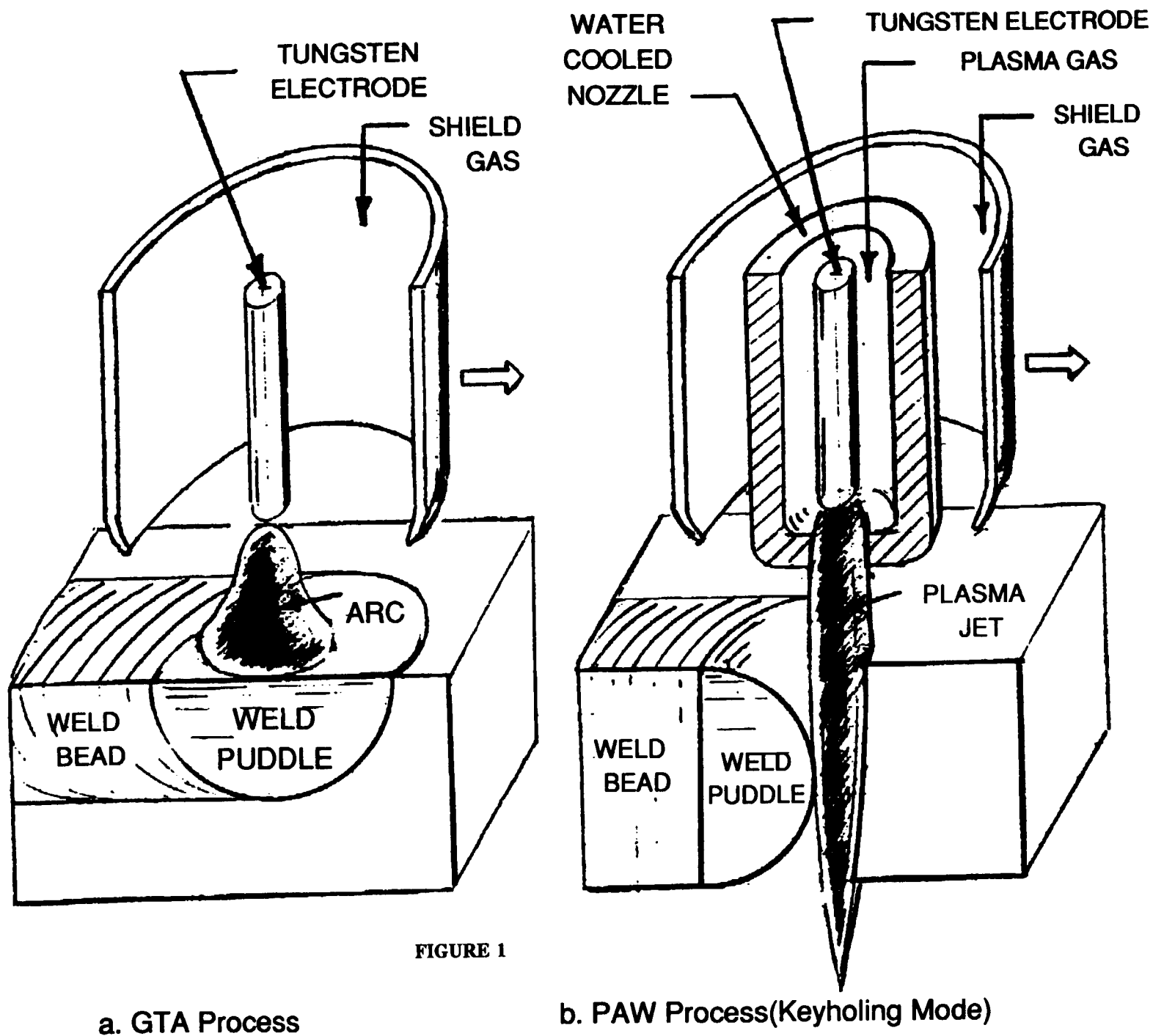
Manufacturers interested in welding heavier sections (say quarter-inch and over) of aluminum alloys or other alloys with tenacious surface oxides have a potential interest in VPPA welding equipment.

For the most stringent aerospace fabrication requirements we regard computer control as essential. When it is ready we think a fully automated, top-of-the-line system will be the most economical for this kind of work.

But it is possible to acquire various versions of a VPPA welding system. A research contractor acquired a rock-bottom basic (strictly manual) system several years ago for less than \$45,000. The equipment was mostly from Hobart with the MSFC special design torch from a local manufacturer (B & B Precision Machine Inc., 6762 Highway 431 S, Brownsboro, Alabama 35741). The machine is used for research, but we believe the (university) contractors can make good welds when they wish even at this primitive level of equipment. Aerospace requirements would not apply in manufacturing, say, tank trucks, for which even primitive VPPA equipment might be adequate.

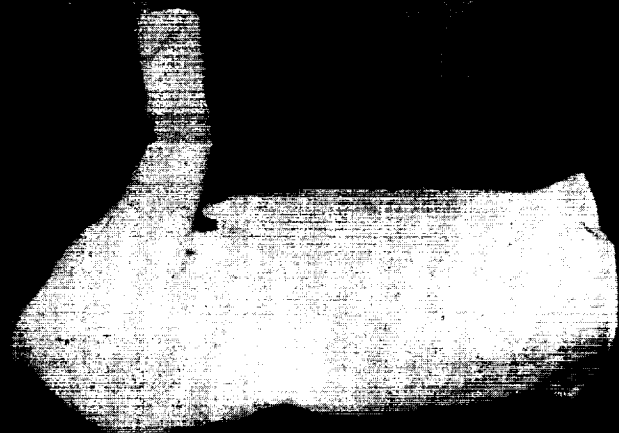
#### **CONCLUSION**

Our intent in making this talk has been to spotlight an item of technology, VPPA welding, which we thought valuable for our own applications and which we chose to, develop further. It is our hope that this presentation may lead to further use and development of VPPA welding to the benefit of a growing technological capacity in the U.S.A. and in the world.



ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH

FIGURE 2





# MARSHALL AUTOMATED WELD SYSTEM (MAWS)

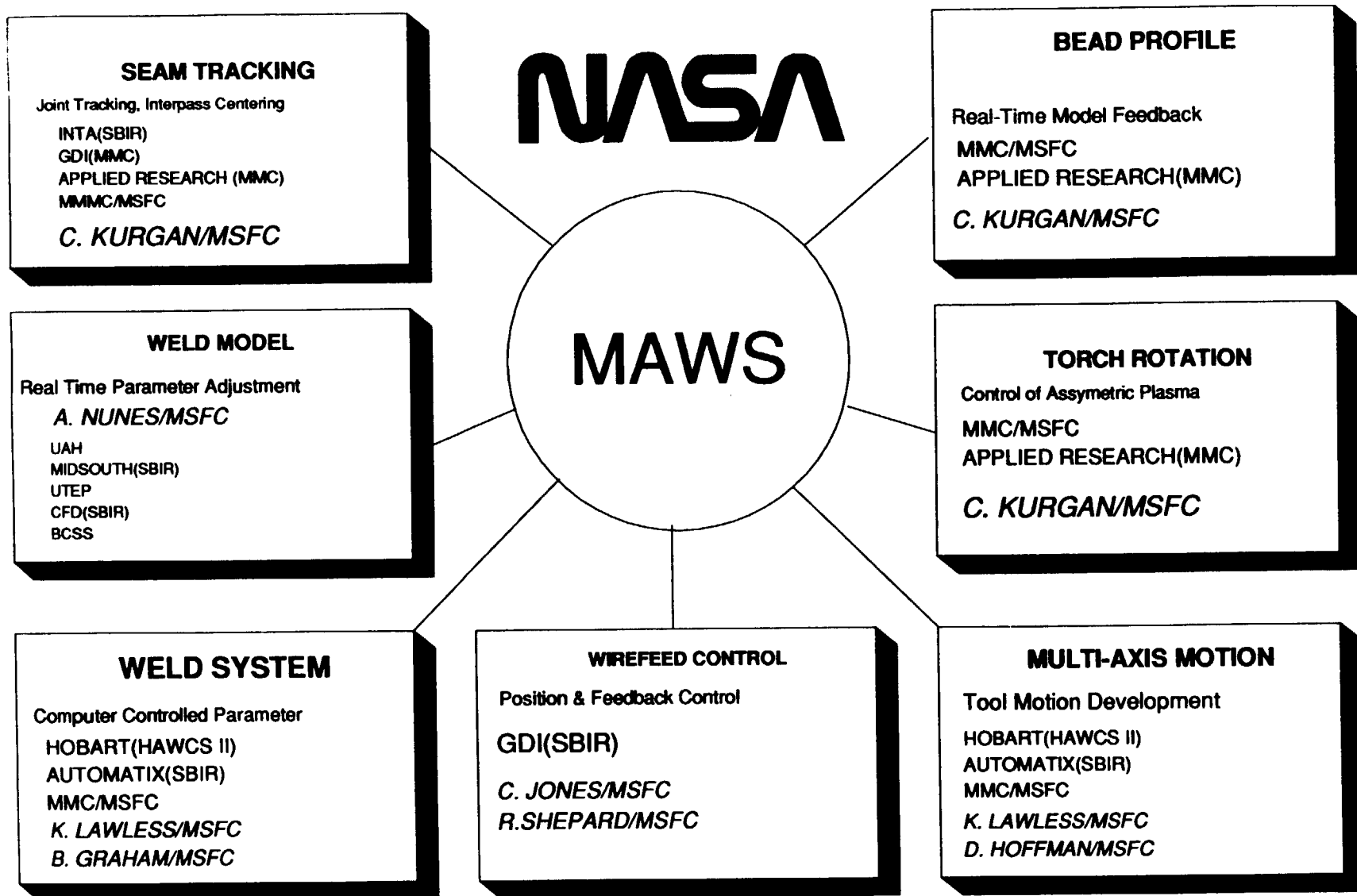


FIGURE 3

**HIGH PRESSURE WATER JET CUTTING AND STRIPPING**

David T. Hoppe  
Majid K. Babai  
Process Engineering Division  
Materials & Process Laboratory  
Marshall Space Flight Center, Alabama 35812

**ABSTRACT**

High pressure water cutting techniques have a wide range of applications to the American space effort. Hydroblasting techniques are commonly used during the refurbishment of the reusable solid rocket motors. The process can be controlled to strip a thermal protective ablator without incurring any damage to the painted surface underneath by using a variation of possible parameters. Hydroblasting is a technique which is easily automated. Automation removes personnel from the hostile environment of the high pressure water. Computer controlled robots can perform the same task in a fraction of the time that would be required by manual operation.

**INTRODUCTION**

Pressurized water cleaning is not a new technology, but it is a process which can always discover new applications. Low pressure water cleaning is an every day occurrence. We wash down driveways with a garden hose, run our cars through magic wand car washes, and rinse plates under the kitchen faucet. Water dissipates a large quantity of energy as it flows which is just waiting to be harnessed. This is seen dramatically in the production of electricity at a hydro electric dam and more subtlety from erosion in our streams and oceans. Pressurized water cleaning utilizes both the force from the impact of water droplets and the frictional erosion force as water flows over the material to cut through and drag particulate matter away. High pressure water cleaning, referred to as hydroblasting, amplifies this power in water to the nth degree. What takes nature years to accomplish, can be performed in minutes at high pressure. A pressure of 10,000 - 15,000 psi is typically used to clean up a solid rocket motor after its retrieval from the ocean. Pressure of 50,000 - 60,000 psi can cleanly cut through the toughest materials.

**BACKGROUND**

The solid rocket motors are retrieved and refurbished after each shuttle flight. With a goal of at least five to six shuttle flights per year and two solid rocket booster motors (SRBs) per shuttle, refurbishment of the SRBs are an important part of keeping NASA's fleet in operation. Hydroblasting is an important and effective step in the refurbishment process. Each SRB is covered with a protective coating of material to protect the base metal from the heat and intense frictional force experienced by the SRBs as they propel the shuttle at incredible speeds, as well as the corrosive ocean environment. These materials referred to as the Thermal Protection System (TPS) must be stripped from the SRB as it is disassembled. Obviously, since TPS is designed to resist these intense frictional forces, it is not easily removed. Once the TPS is removed, the painted surface is inspected. If the top coat has been damaged it too is stripped off. The base coat can also then be stripped if necessary. The variation of several parameters control the stripping effectiveness of the water spray. The angle of incidence of the water to the SRB, water pressure, flow rate, type of nozzle, distance between nozzle and SRB, water jet traversing speed, and direction

of water jet traverse all have an effect on the process and must be optimized to produce the desired stripping results.

These parameters cannot be effectively controlled by manual operation which results in excessive processing time due to the reduced efficiency. Manual operation of a hydroblast system is also a dangerous and difficult operation. The effect of the strain of manual operation, can be compared to the mess generated while washing a car combined with the noise and fatigue from operating a jack hammer. Controlling a water stream at a pressure that can slice through the TPS places the operator in constant peril of being cut himself by the water spray. An operator must wear full body protection and a face shield equipped with an air hose. The operator must resist a 75 pound backthrust which is produced from a water spray pressure of 10,000 psi. To remove the operator from the direct contact with the water spray, the hydroblast process has been automated with the aid of a computerized robot. Operators can then set behind the console of a computer in an airconditioned and sound proofed control room. The hydroblast system setup for the removal of TPS from the SRB aft skirt is shown in Figure 1.

### **AUTOMATED SYSTEM OPERATION**

The hydroblast refurbishment facility installed in hanger AF at Kennedy Space Center was developed in the Productivity Enhancement Facility at the Marshall Space Flight Center (MSFC). The hydroblast system designed at MSFC utilizes a frequency control Hammelmann pump system. A water diversion system switches the water flow between different nozzles without physically removing and installing the nozzles. The nozzles to be used are attached to the end of a six axis of freedom, hydraulic operated robot which is set on a lift table. The robot arm can resist 225 pounds of reaction force and maintain an accuracy of 50 mils. This allows full access to an entire section of a component of the SRB directly in front of the robot. The component sets on a turntable which rotates the component in front of the robot allowing full coverage of the component by the robot. The turntable has an angular positioning accuracy of one quarter of a degree. A DEC PDP 11/23 computer controls and synchronizes the robot program, pump pressure and water flow rate, and the turntable location and speed. The hydroblast system components are shown in Figure 2.

Parameter optimization must be performed for each material the hydroblast is used to remove. TPS materials which have effectively been removed by the hydroblast include specially formulated Marshall Sprayable Ablatives (MSA), K5NA, Marshall Trowellable Ablator (MTA), PRC - 1422, and Cork. Gloss white Bostick Epoxy paint #443-3-1, different types of adhesives, and even three inch thick insulation. Although each material has its own characteristics, some general principles can be applied to any material however. Stripping material from the top down reduces the interference with the water stream by the material as it is removed. It is best to have the angle of incidence directed toward the material to be removed. After the water jet cuts through the material, it penetrates between the metal surface and the material to be removed which aids the stripping of the next pass. A glancing angle of incidence was found to be best for stripping TPS without damaging the topcoat of paint. A near perpendicular angle of incidence was more effective at removing the white Bostick top coat.

At a 30 degree angle of incidence, a twenty four inch water jet length between the nozzle and the SRB component was found to be optimal for removal of TPS. This physically placed the nozzle within twelve inches of the SRB. At a twenty-four inch water jet length, the water spray pattern impacts the component before it can dissipate. A water jet length less than twenty-four inches results in a more concentrated impact pattern and although this would yield a stronger stripping force, the smaller removal area results in less efficient stripping. A water jet length of eighteen inches proved effective for the removal of the top coat of paint. Conventional nozzles proved to be inadequate for the high pressures associated with the hydroblast system. These nozzles showed low stripping performance and high erosion degradation due to inefficient flow characteristics. The available nozzles

were rated at 10,000 psi or lower. High Pressure Full Tapered Stripping (FTS) nozzles were designed with a long convergence before the exit orifice in order to raise the performance to an acceptable level. These nozzles were also manufactured from a tougher steel. The operation of the automated hydroblast robot has resulted in an estimated savings of \$52,000 per launch.

### **PROCESS ENHANCEMENT AT MSFC**

An additional process enhancement investigated was the use of a rotating nozzle. A twin jet nozzle fixed on a rotating hub is turned by the back pressure of the water jet. Moving the position of magnets within the housing changes the amount of resisting force to provide speed control over the nozzle rotation. At 15,000 psi, a rotational speed of 1500 rpm is possible. This rotating nozzle removes a much wider strip across the material at each pass. Test results show this to be effective at removing TPS type material but not paint.

Dry grit blasting is a common stripping process and is generally required to remove paints, especially primer coats. Dry blasting methods used to remove paint and primers greatly increase the time and cost for processing. Combining grit blasting with hydroblasting will greatly reduce the refurbishment time and movement and handling of hardware required. A mixing chamber is added to the nozzle assembly. A gravity feed hopper is connected to the nozzle mixing chamber where the vacuum created by the water flow through the nozzle pulls in the grit material. Crushed walnut shells have proven to be an excellent grit material for the removal of primer paint. The crushed walnut shells are soft enough to prevent impact damage to most base materials.

Caution and testing must be performed before any production usage of the hydroblast can be performed. The high pressure water spray is a potent force. At 15,000 psi, a water spray that is allowed to dwell at the same point on an aluminum plate will cut into the plate and if left long enough would eventually cut all the way through. It is a good habit to start the robot or turntable motion before the pressure is increased or to increase the pressure while the spray is being dissipated into the air surrounding the part to be hydroblasted in order to preclude this possibility. Several feet is all that is required to spread the water blast spray out into turbulence and quickly dissipate its energy.

The extremely high pressurized water spray is used for cutting many materials. The walnut shell grit injection is used with pressures up to 60,000 psi to accomplish an extremely effective cutting process. The high pressure produces a flow rate at three times the speed of sound. Special sapphire nozzles must be used because of the high erosion characteristics of this flow. Even though it cuts fairly cleanly, final milling of the part would still be desirable. The ultra high pressure water cutting referred to as the water knife can reach and effectively cut through parts where conventional cutting techniques cannot reach. The water knife has been coupled to a tracer mill to quickly cut out patterns in hard materials such as inconel in a fraction of the time usually required to rough cut these materials. It is effective on composite materials slicing through the fibers without causing fraying.

### **MOBILE ROBOT**

This automated robot is great for regular and repeatable operations, but some of the SRB refurbishment requirements are not so easily isolated or have easy access. Even after several years of use there have been areas on the SRBs that required manual attention, until now. Marshall Space Flight Center (MSFC) has recently developed a mobile robot to eliminate these last manual hydroblast requirements. The Mobile Robotic Hydrolasing System (MRHS) utilizes a six degree of freedom articulated robot mounted on a transportable platform. The vehicle can be raised up to 34 inches off the ground by four

hydraulic cylinders for extra reach towards the tallest points on the SRB. Air conditioning and soundproofing were added to the cab's new watertight plexiglass cabin, creating an enclosure that will protect the operators and robot's controller from heat, humidity and debris. The robot is electric but has similar capabilities to the hydraulic robot. The MRHS includes similar diversion valve switching between nozzles. Also included were an inclinometer for deck leveling and a displacement transducer to input the deck height to the robot's controller for verification and added safety.

This unit, which is first of its kind, can be maneuvered around the boosters. The robot's end effector holding the nozzles can withstand the delivery of 25 gallons of water per minute under 15,000 psi pressure. The nozzle assembly moves through a smooth preprogrammed robotic path removing TPS from the boosters. This is twice the flow rate of water which can be handled manually, with a consistent repeatability of 0.02 inch accuracy. It improves quality of the TPS stripping and increases personnel and hardware safety while reducing refurbishment time. It has been conservatively estimated that the system can save 20 hours of serial time per flight set. This equates to a 40 percent reduction in TPS removal time prior to SRB disassembly at an estimated savings of \$67,200 per launch. As to its untapped potentials, what the Mobile Robot offers is a prototype for a traveling, ergonomic control center with attached robot that could expel mediums other than water. Already under evaluation at MSFC and USBI is its feasibility for sealant application, paint spraying, and wet ablator insulation. This system has applications yet to be imagined for other NASA programs, for the military and commercial aerospace. The conceptual design for the mobile robot is shown in Figure 3.

### CONCLUSION

Hydroblast stripping procedures have proven to be an important part of the refurbishment of solid rocket motors. The fully integrated computer controlled hydroblast system provides a safe and efficient facility for the refurbishment cycle. The development of the Automated Water Blast Research Cell at the Productivity Enhancement Facility has resulted in the savings of millions of dollars and hours of precious processing time at Kennedy Space Center. The newly developed mobile robot has further increased the savings in cost and time. The technology developed for the mobile robot can be used for new applications of the mobile robot concept. New usage for the hydroblast system are always being discovered for hard to remove materials. The hydroblast system has proven to be one of the most reliable and adroit systems at both the Marshall Space Flight Center and the Kennedy Space Flight Center.

### ACKNOWLEDGEMENTS

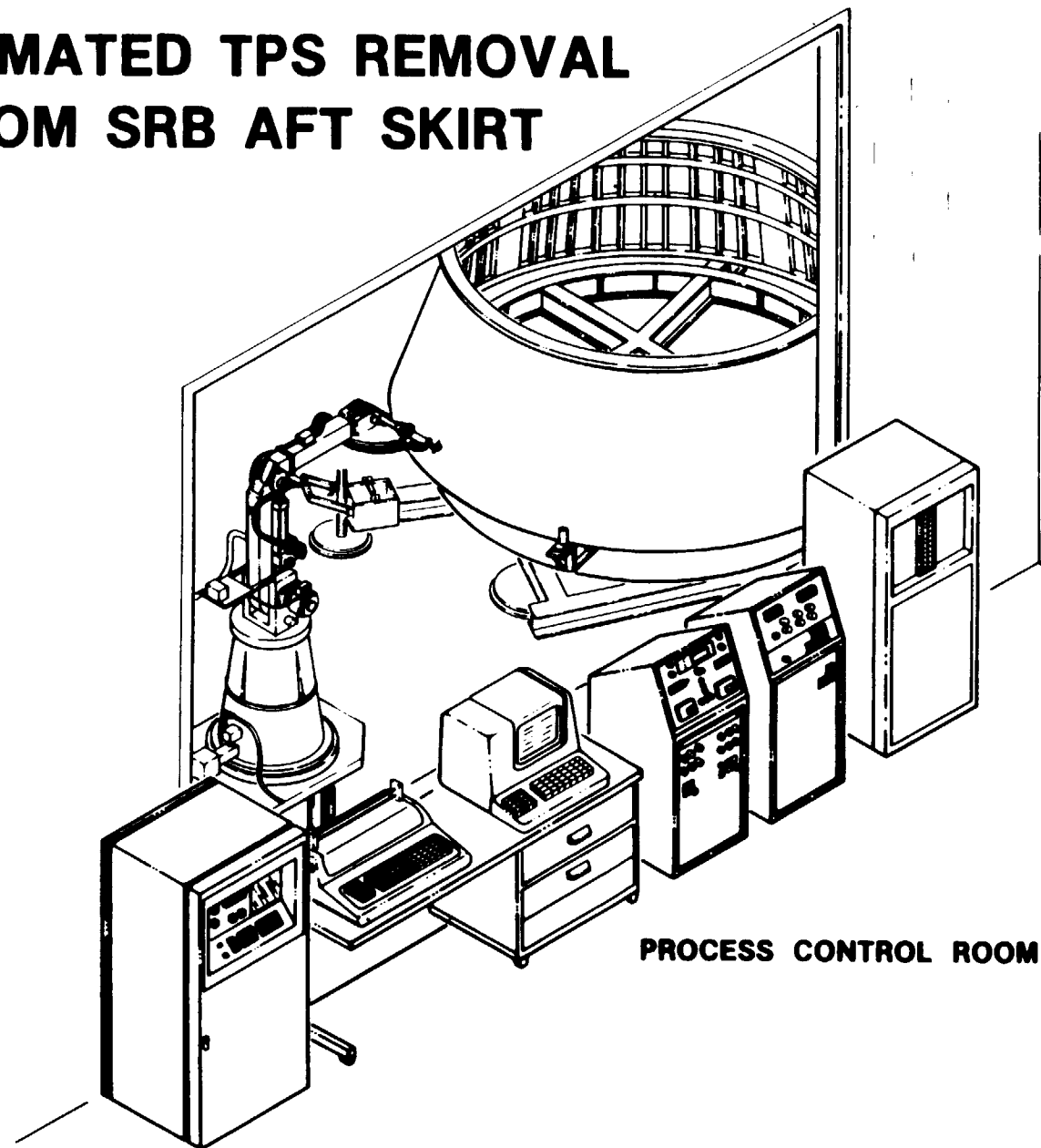
This project was initiated by Marion Roberts, NASA retired, who with his associates designed and installed the initial system at Marshall Space Flight Center, and developed the FTS nozzle.

Robert Rice, Steve Cosby, and Fletcher Burgess and associates of USBI have been instrumental in the installation, operation and optimization of the hydroblast process.

### REFERENCES

Robert M. Rice, "Process Report on the Automated Hydro Removal of TPS" prepared for NASA, Contract NAS8-36300, January 1986.

# **AUTOMATED TPS REMOVAL FROM SRB AFT SKIRT**



**FIGURE 1**

# SRB AUTOMATED TPS REMOVAL SYSTEM

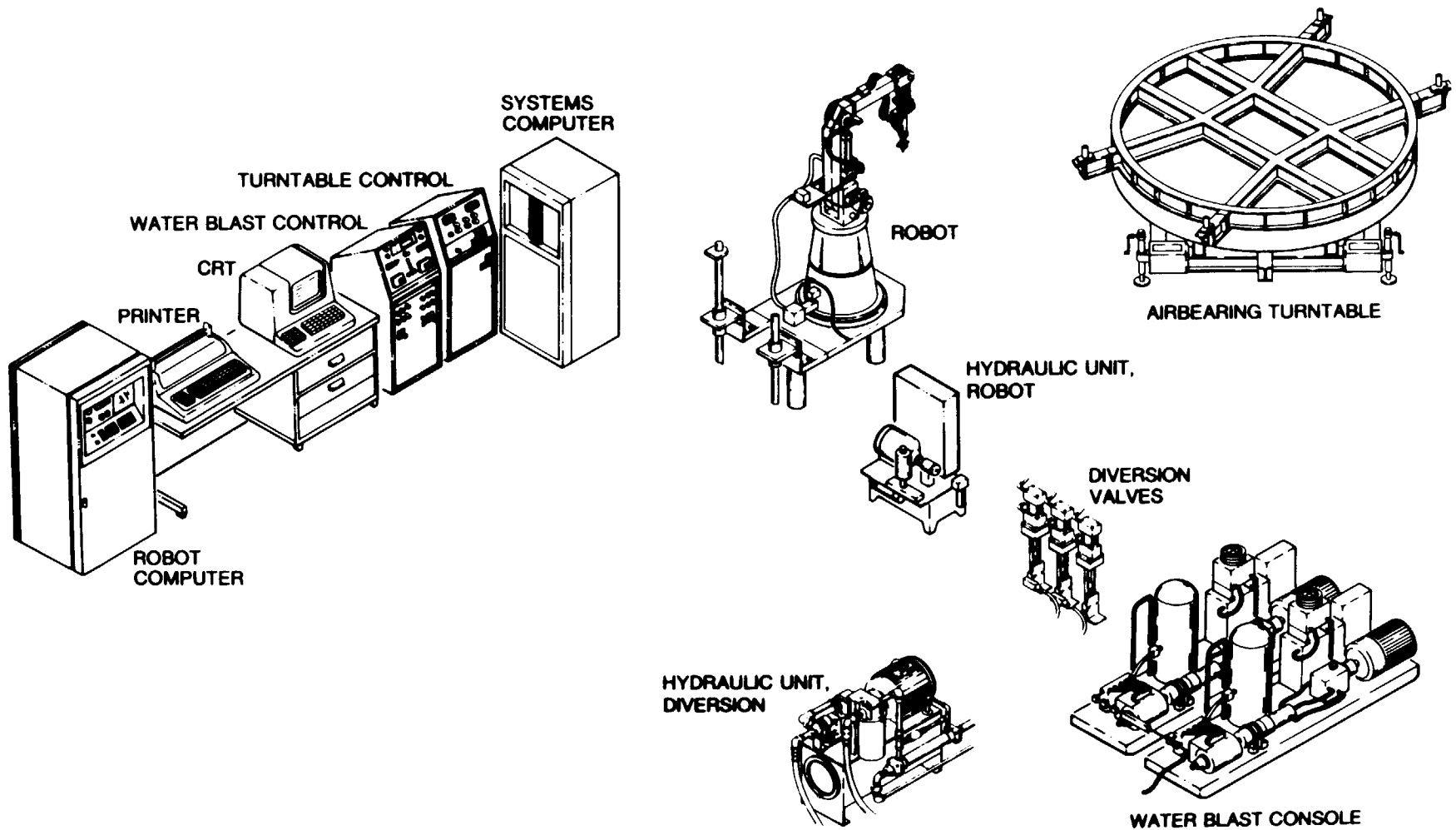
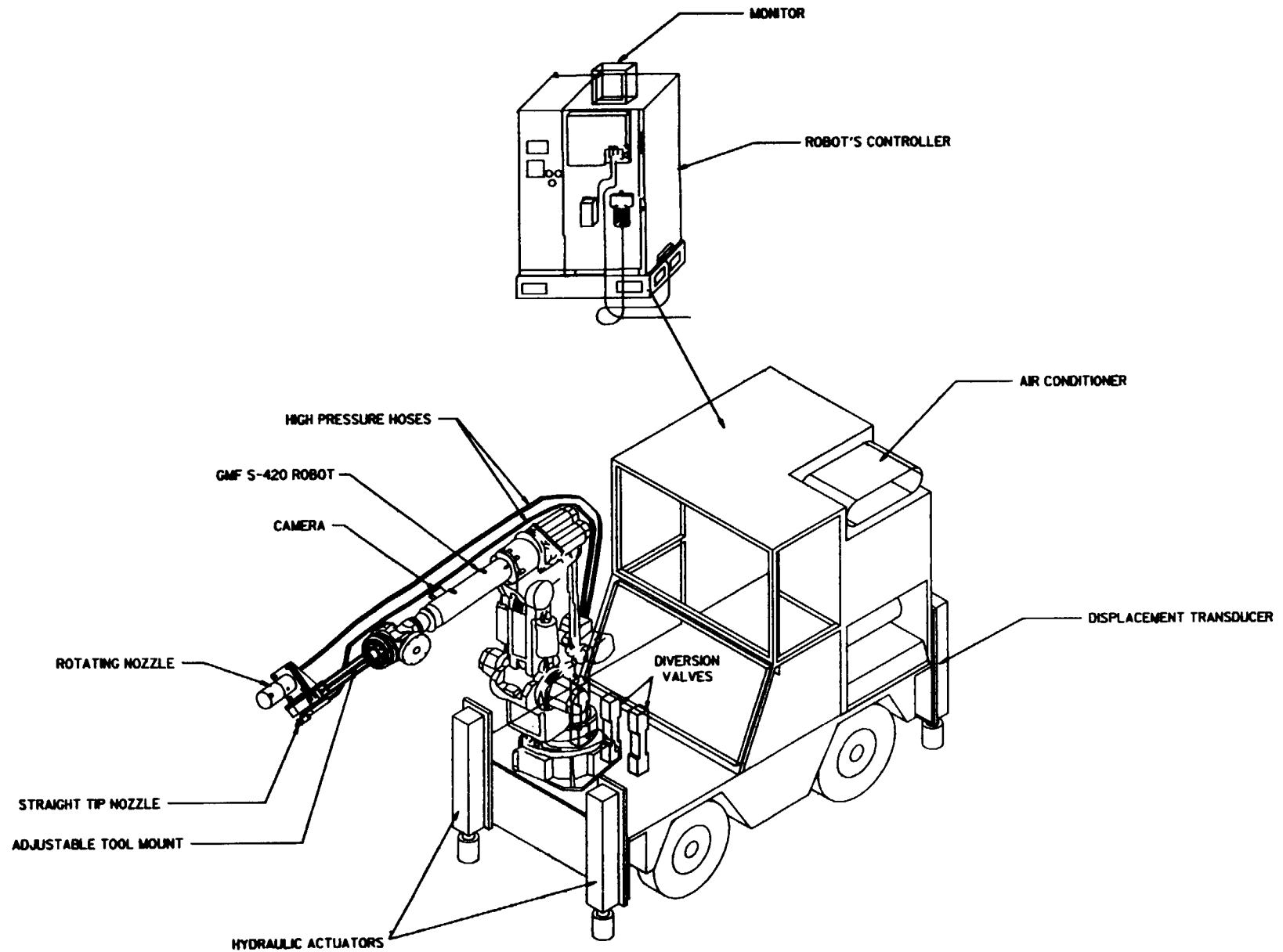


FIGURE 2

# MOBILE ROBOTIC HYDROLASING SYSTEM





## Cost-Efficient Manufacturing of Composite Structures

W. Tom Freeman, Aerospace Technologist-Structures Technology Program Office  
 Dr. John G. Davis, Head-Structures Technology Program Office  
 Dr. Norman J. Johnston, Chief Scientist-Materials Division  
 NASA Langley Research Center  
 Hampton, Va.

### Abstract

The Advanced Composites Technology (ACT) program is seeking research breakthroughs that will allow structures made of graphite epoxy composite materials to replace metal in the wings and fuselages of future aircraft. NASA's goals are to reduce acquisition cost by 20-25 percent, structural weight for a resized aircraft by 40-50 percent, and the number of parts by half compared to current production aluminum aircraft. This presentation will focus on the innovative structural concepts, materials, and fabrication techniques emerging from the ACT program, and will discuss the relationship between aerospace developments and industrial, commercial, and sporting goods applications.

### Introduction

Boeing Commercial Airplane (BCA) Group and Douglas Aircraft Corporation (DAC) use approximately 400,000 pounds of composites per year in spoilers, rudders, elevators, doors, and other secondary structure. The rate of application of composites to empennage, wing, and fuselage commercial airframe primary structure has been disappointingly slow. Composite materials are an obvious choice for performance optimization, corrosion resistance, and fatigue suppression but before a bold leap toward more extensive use of composites can be expected in commercial applications, lower acquisition costs and confidence that production costs can be accurately predicted must be demonstrated. The Advanced Composite Technology Program's goal is to establish the design concepts, develop manufacturing approaches, and demonstrate mechanics and cost performance of innovative low cost composite assemblies providing confidence for production commitment to primary structure by the turn of the century.

### State-of-the-Art

The application of composite structures to commercial transport aircraft is proving to be an evolutionary process. The National Aeronautics and Space Administration (NASA) has sponsored the development, certification and flight service evaluation of composite secondary and empennage structure for components shown in Table 1 (1). Secondary and empennage composite structures are in production on

Component	Secondary			Empennage		
	DC-10 rudder	727 elevator	L-1011 aileron	DC-10 vertical stabilizer	737 horizontal stabilizer	L-1011 vertical fin
Size (root X span), ft.	3.2 X 13.2	3.4 X 17.4	4.3 X 7.7	6.8 X 22.8	4.3 X 16.7	8.9 X 25
Metal design weight, lb.	91	130	140	1,005	262	858
Composite design weight, lb.	67	98	107	834	204	622
Weight reduction, %	26	25	24	17	22	28
Number of production units	20	11	12	3	11	2
Start flight service	4/76	3/80	3/82	1/87	3/84	-

Table 1. - Composite structures developed under NASA ACEE program.

several transport aircraft. The weight reduction potential of composite structures is well documented in an excellent bibliography with over 600 references on NASA sponsored composites research between 1976 and 1986 that was published in 1987 (2). Boeing 737 spoilers for the NASA flight service program have been in service since 1973 and after 15 years of flight service indicate no loss in residual strength (3).

Flight service experience made possible the production commitment of secondary composite structures on the Boeing 767 shown in Figure 1. Full exploitation of the benefits of composites requires application to wing and fuselage structures which account for 75 percent of the aircraft structural weight. The cost to develop and produce composite structures remains the major barrier to increased application of this technology to transport aircraft, shipbuilding and general industrial use.

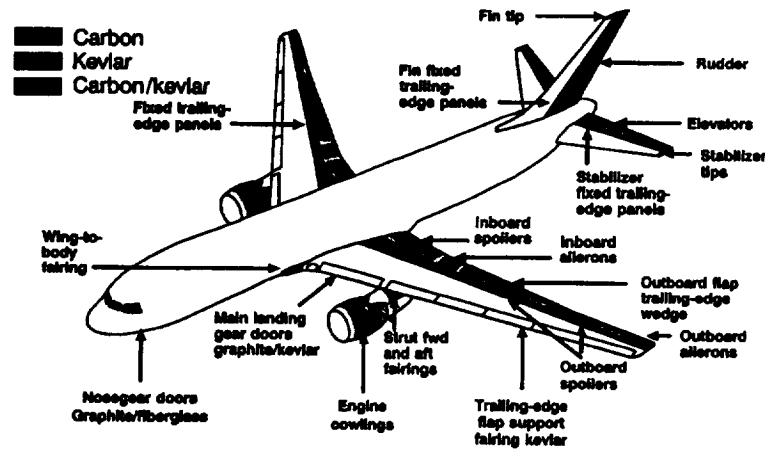


Figure 1. - Current Boeing Commercial Airplanes composite usage.

### The ACT program

In 1984 NASA requested that the Aeronautics and Space Engineering Board of the National Research Council (NRC) form a committee to assess the status and viability of organic matrix composites technology for aircraft structures. The committee findings and recommendations were reported in 1987 (4) and recommended that NASA establish a program to develop an integrated "affordable" composites technology database to reduce the risks of employing new cost effective designs and fabrication processes. In 1988 NASA initiated the ACT program with emphasis on development of advanced materials, mechanics, innovative concepts, and low cost manufacturing methods. This paper will provide detail on the improved material forms, tailored cost effective design concepts, and innovative low cost manufacturing methods that are under development and expected to lead to an "integrated database" of affordable composite concepts. The term "integrated database" relates to materials and structural issues required to reduce the risks of applying new composite designs to large structure production applications. Data to be developed include materials properties, test methods, analytical methods for predicting structural response and strength, manufacturing and inspection methods.

Until the database for the composite structures reaches a level of maturity comparable to that for current metallic structures, fabrication and testing of full-scale components will be required to demonstrate performance and cost goals. Before the commitment to develop and build a composite primary airframe is made, full-scale fabrication and tests of a wing box, a fuselage barrel and possible the wing-fuselage attachment will be required to demonstrate that an "affordable" database exists. Technology verification for the next decade is expected to require fabrication and testing of full-scale wing-box and fuselage-section components before certification can occur and production commitments can be made.

### Potential Benefits

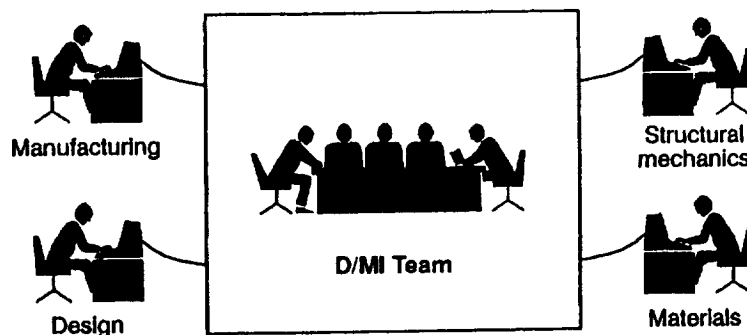
Composites structures are recognized as enabling technology for making significant advances in the performance of subsonic transports and are essential to achieving an economically viable supersonic transport. Prior research and development programs have demonstrated weight savings greater than 25 percent and, when the effects of improved damage tolerant materials and resizing the aircraft are taken into account, the weight savings for some components may approach 40 percent. Composite structures that have been designed as replacements for metallic structures contain significantly fewer parts and numerous studies have noted the direct relation between part count and cost. Cost estimates that range from 10 to 40 percent less than metallic structure have been reported for the projected acquisition cost for composite

structure in future production (5). The weight savings of composite structure translates into about 10 to 15 percent fuel savings (6). Outstanding resistance to environmental degradation and residual strength after exposure to cyclic fatigue loading offer the potential for building aircraft with longer life and reduced maintenance cost (7). Recent studies indicate that composite structures are essential for design of an economically-viable 5000-nautical-mile-range supersonic aircraft (8). Composites are required to hold the gross takeoff weight to less than 750,000 pounds and allow landing on existing runways.

The recent emphasis on reducing composite structure acquisition cost to less than equivalent metallic structure cost will provide a broad range of opportunities for application of high performance composite materials in shipbuilding and heavy industry with renewed opportunities in the automotive, commercial, and sporting business sectors. Order of magnitude increases in volume requirements for composites can be envisioned in the next 2 decades as commercial airplane wing and fuselage and large shipbuilding applications evolve, resulting in additional decreases in raw material cost. New indefinite room temperature storage composite material textile forms are expected to make warehousing of structural building blocks possible and high modulus graphite composites routinely available to the home hobbyist and building construction markets.

### Design/Manufacturing Engineering

The preliminary design phase has been identified as the most opportune period during an airframers hardware production cycle for substantial cost reduction. Boeing has experienced that 70% of airplane fabrication costs are fixed by the time the design is frozen and that the influence of engineering on fabrication cost reductions is significantly reduced once the design is completed. Concurrent engineering interdisciplinary teams are emphasizing cost evaluation at the early stages of the development cycle. Perhaps, the most important approach for achieving an affordable composites data base is through the integration of design and manufacturing disciplines. The Design/Manufacturing Integration (D/MI) approach brings together all disciplines (see Fig. 2) that have an influence on the performance and cost requirements



**Figure 2. - Design/manufacturing integration (D/MI) program.**

at the beginning of the design process. Figure 3 provides a schematic of the results of the DM/I process for a simple stiffened skin compression panel. The influence of design concept on cost and weight is the product of the D/MI process. The interaction of interdisciplinary teams can perform trade studies quickly resulting in less changes as the design is completed. Airframe companies have begun to practice the D/MI approach and large benefits from this approach will appear in the next generation of aircraft.

Rapid advancements in computer hardware and software capability at continually reduced cost has made computer aided design and manufacturing (CAD/CAM) practical for storing complete 3-D electronic product definitions of complex composite structures. These concepts proven on the B-2 program "minimize costly and time consuming toolmaking phases" normally associated with development of composite structures (9). Companies employing computers to rapidly manipulate 3-D hardware definition databases will develop D/MI capabilities that allow the designer to optimize composite material properties while tailoring the performance and cost of the structure. One ACT program goal is to develop a designers cost model that will be integrated with a structures optimization model to trade cost and weight during the conceptual and preliminary design phases.

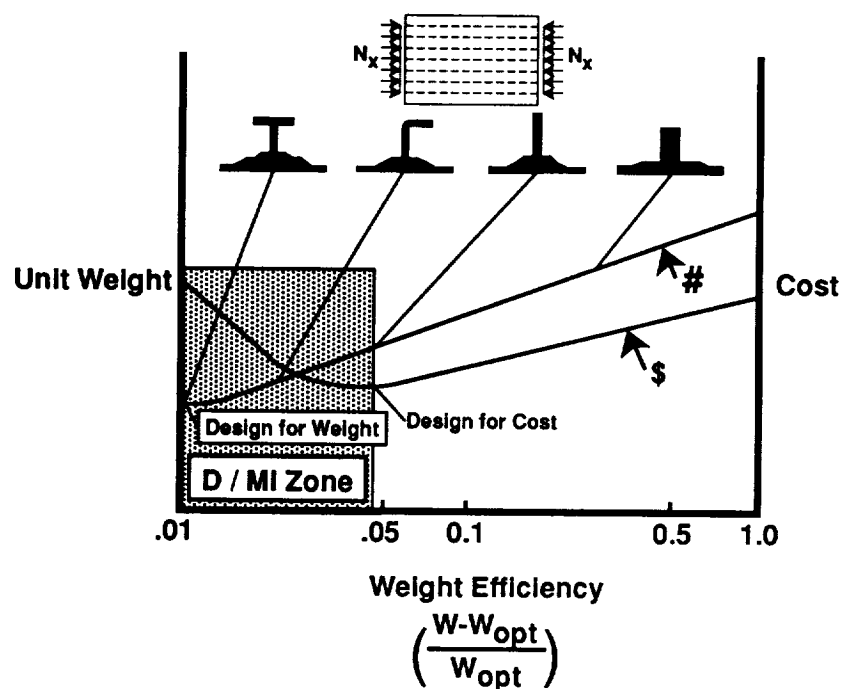


Figure 3. - Influence of design/manufacturing integration.

#### Design Concepts

Composite structures for aircraft applications have historically been designed using standard metallic stiffened-skin design practices. Composite design concepts are now being developed to reduce both weight and cost. Design concepts that reduce fabrication and assembly costs can often be cost-effective even with more expensive materials. Boeing is studying pressurized fuselage concepts for the ACT program using design concepts that result in lower cost structures by reducing the number of parts and associated fasteners that have to be assembled. Figure 4 shows a concept to build large one piece composite panels reducing the total number of panels for a 757 class airframe to 19 panels. The equivalent

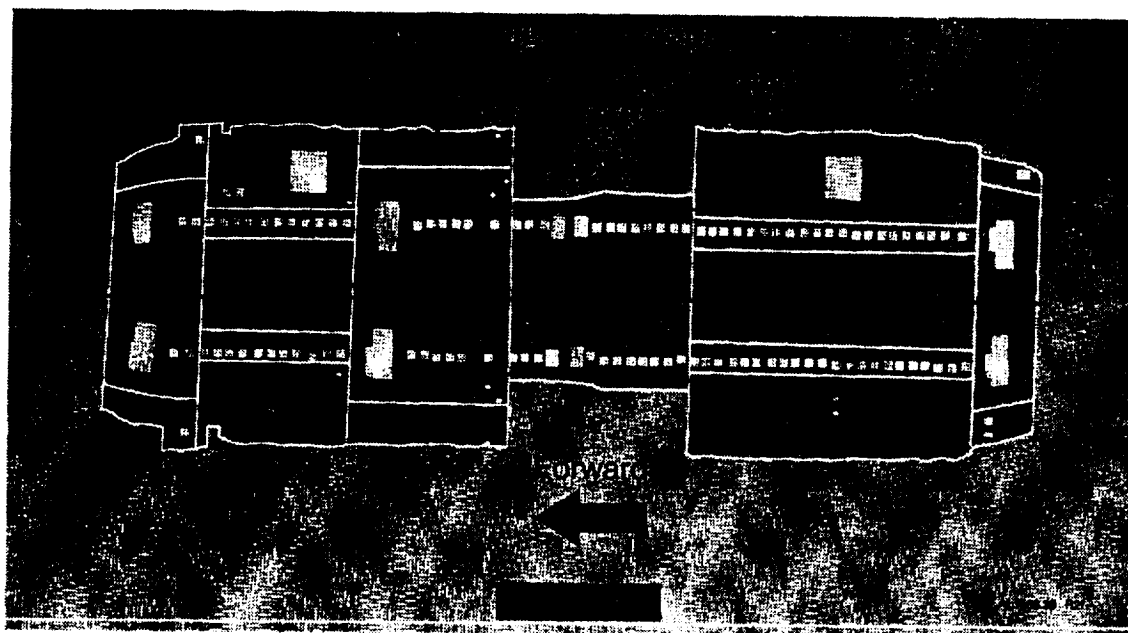


Figure 4. - Design concepts

ORIGINAL PAGE IS  
OF POOR QUALITY

metallic structure has 72 panels due to limits on sheet metal width availability. Several structural elements can be integrated into the composite structure during subcomponent fabrication by cocuring the stiffeners with the skin to significantly reduce the number of fasteners, clips and shear ties needed to assemble conventional metallic stiffened structures. Sandwich structures are being considered to minimize or even eliminate the need for stiffeners. Figure 5 shows concepts for composite fuselage quadrant panels designed

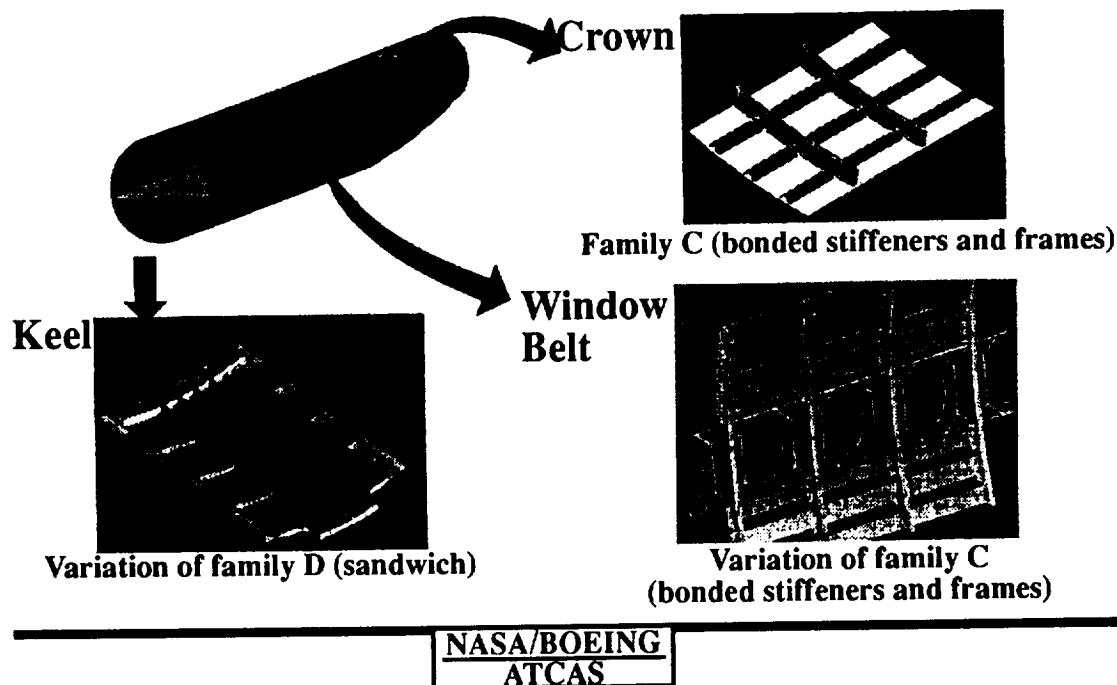


Figure 5. - Baseline fuselage concepts.

for an all composite aircraft rather than a conventional metallic replacement. Large one piece composite panels (up to 32' by 20') can eliminate dozens of complex joints and countless thousands of fasteners that are common to metallic assemblies. Boeing estimates as much as a 51% weight savings with acquisition costs equal to equivalent metallic structure for composites built by these concepts, even with today's high cost materials. Their confidence that large one piece structures can be fabricated comes from experience noted in an Aviation Week article, "Boeing's B-2 facility Produces Largest Composite Structural Parts Ever Made" (10).

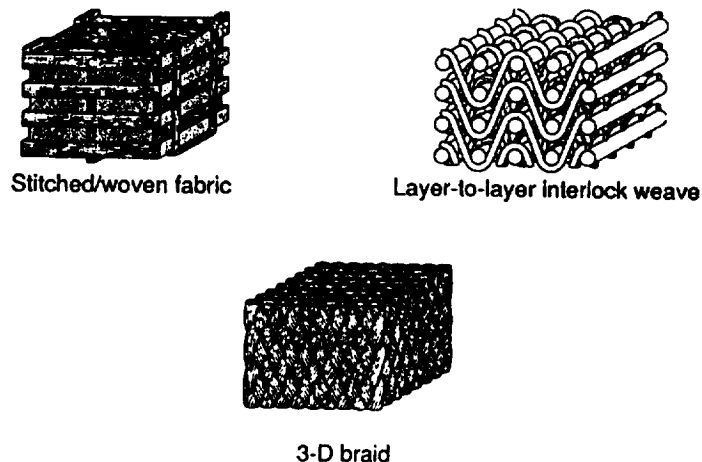
#### Materials and Material Forms

To successfully replace state-of-the-art (SOTA) metallic materials, carbon fiber reinforced polymer composites must exhibit a favorable balance between toughness, damage tolerance, processability, and cost. The primary resin matrix for subsonic aircraft structure in service today is a brittle 350°F curing epoxy system that does not exhibit the toughness and damage tolerance required for heavily-loaded primary aircraft structure. Toughened matrix materials now cost at least three times that of the conventional brittle epoxy systems and are generally too expensive for widespread use on commercial aircraft (11).

Thermoplastic composites offer an attractive combination of mechanical properties, toughness, and the potential for low-cost manufacturing. They provide simple formulations, excellent prepreg stability, indefinite work-life, no requirement for cold storage, and rapid thermoformable processing. Solvent resistant thermoplastic tape is currently much more expensive than conventional brittle epoxy prepreg. ACT Program participants are developing and evaluating powder coated fiber, textile preform technology, and resin transfer molding (RTM) materials to reduce the materials cost associated with composites.

ORIGINAL PAGE IS  
OF POOR QUALITY

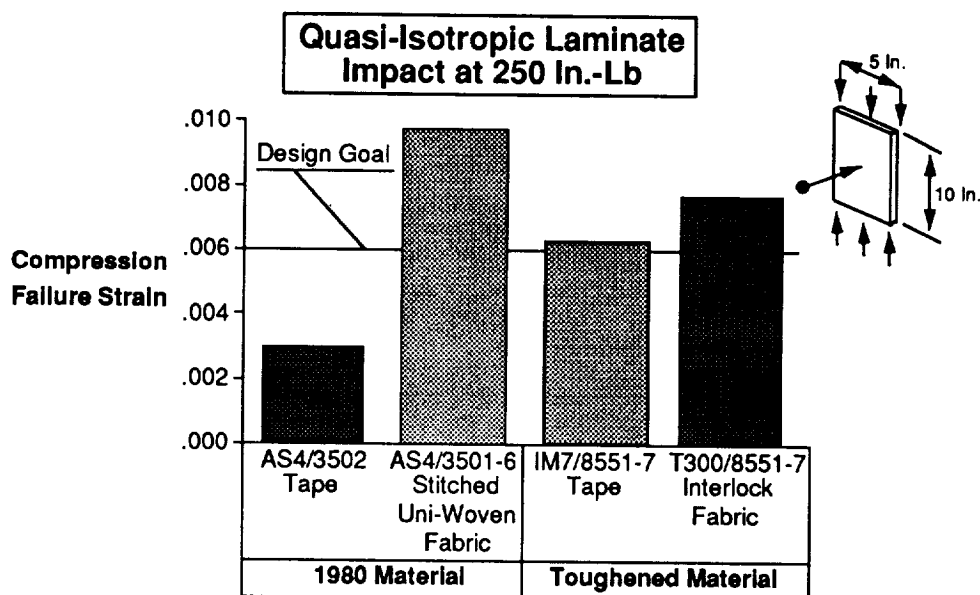
An exciting approach to low cost composites that exhibit exceptional toughness is the use of net-shaped textile fiber architecture preforms that can be infiltrated by RTM processes and cured in an oven or autoclave (12,13,14). Automated textile processes such as weaving, braiding, knitting and through-the-thickness stitching would provide highly automated processes for reduced material cost. Complex fiber architectures as shown in Figure 6, multilayer fabrics and structural composite preforms have been demon-



**Figure 6. - Through the thickness reinforced textile forms.**

strated using automated textile technologies. These dry fiber preforms lend themselves to economical composite fabrication techniques including RTM and pultrusion. Stitched fabric preforms impregnated with low cost epoxy matrix resins exhibit outstanding damage tolerance and out-of-plane load capability.

The compression after impact test results for a stitched uni-woven fabric flat panel made with conventional brittle epoxy resin shown in Figure 7 exhibit a failure strain far above the accepted design



**Figure 7. - Standard compression after impact for quasi-isotropic laminates, impact at 250 in.-lb.**

goal of 0.006% (11). Stitching provides the lowest cost highest productivity approach for through-the-thickness reinforcement of low cost 2-D fabric preforms. These preforms can be stored indefinitely at ambient conditions and resin transfer molded on demand with low cost two part epoxy matrix systems.

Douglas Aircraft Corp. is developing the technique for constructing stitched dry preforms from uni-weave fabric subelements illustrated in Figure 8 (15,16). The matrix used in the Douglas process is

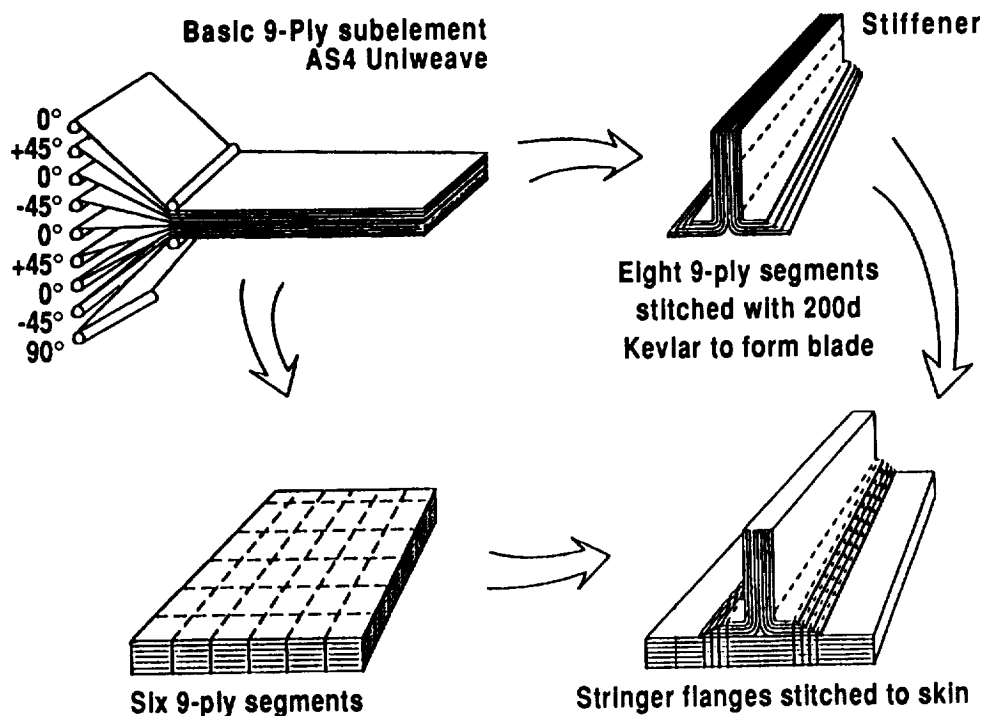


Figure 8. - Damage-tolerant stiffened panel concept.

either vacuum or pressure infiltrated into the preform which is bagged and cured using standard autoclave procedures. Damage tolerant single and triple blade-stringer panels have been fabricated from dry stitched preforms with conventional epoxy RTM. A 22-inch-wide by 22-inch-long three-stringer panel demonstrated excellent post-impact compressive strength as shown in Figure 9 (15).

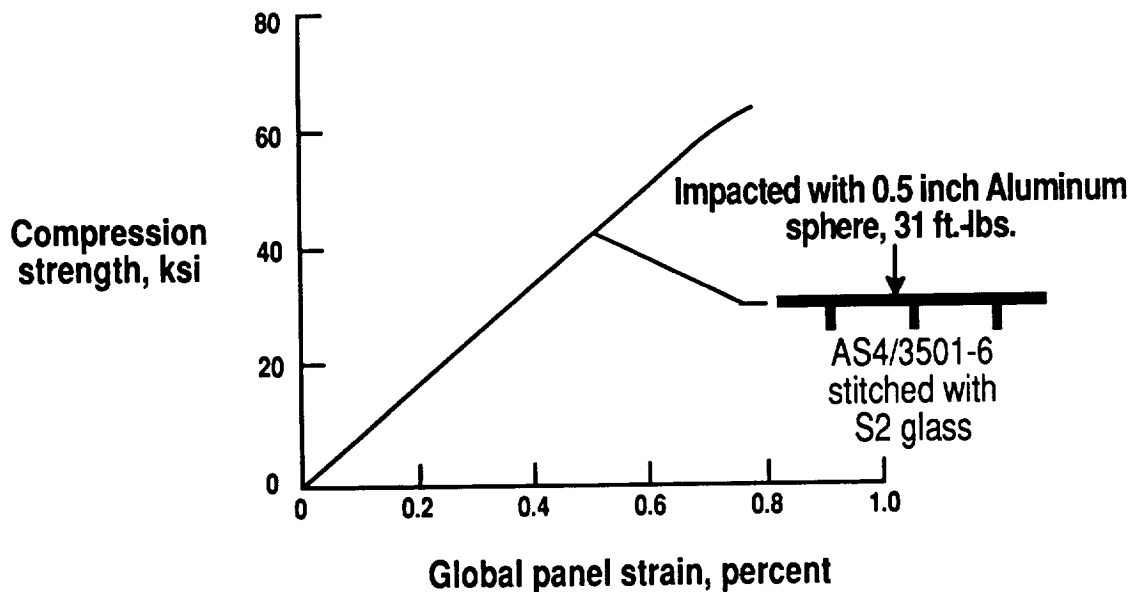
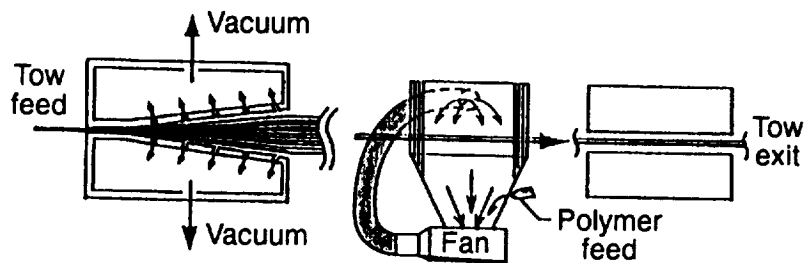


Figure 9. - Post impact compression strength of blade-stiffened panel.

Alternate methods to impregnate fibers with resin while maintaining the textile preform concept are being developed. Powder coating processes that homogeneously apply a polymer powder to 3K to 12K carbon tow offer promise as low cost towpreg that can be woven into fabric. The coating process is relatively simple and involves a tow spread chamber, a powder recirculating chamber, and a convection



## Prepreg Processing Modules

Tow spreader chamber	Powder coater	Powder fusion
<ul style="list-style-type: none"> <li>• Adjustable tow spread</li> <li>• Minimum fiber damage</li> <li>• Fast line speed</li> </ul>	<ul style="list-style-type: none"> <li>• Recirculating chamber</li> <li>• Uniform coating</li> <li>• Precise resin control</li> </ul>	<ul style="list-style-type: none"> <li>• Convection furnace</li> <li>• Individual fiber coating</li> </ul>

Figure 10. - Dry powder prepping.

furnace for powder fusion as shown in Figure 10 (17). Advantages of powder-coated towpreg as a new method to combine fiber with resin are given in Table 2. It is especially useful for thermoplastics which

- 
- Versatile: thermoplastics and thermosets
  - Room temperature
  - No solvents involved
  - Minimum exposure to toxic materials
  - Prepreg requires no refrigeration: cuts waste/spoilage
  - Prepreg can be woven, filament wound, pultruded, thermoformed
  - Good alternative to RTM processing of textile preform composites
- 

Table 2. - Importance of the powder coating process.

otherwise are difficult to prepreg onto carbon fiber without the use of difficult hot/melt or solvent coating procedures. The mechanical properties of unidirectional and woven composites made from powder-coated towpreg show promising results (18).

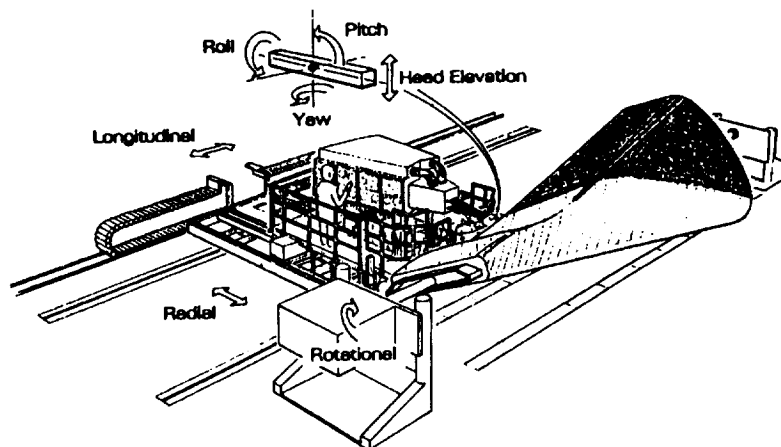
## Manufacturing Methods

ACT program participants are developing lower cost starting materials that do not add to down stream manufacturing cost, significantly more automation that will reduce labor cost and eliminate scrap due to human error, simple designs that exploit automated processes, approaches to eliminate autoclave consolidation requirements and assembly processes that eliminate shimming. Specific processes and methods that are being investigated include building block textile preforms, resin transfer molding, advanced fiber placement and braiding. Through-the-thickness stitching of uniweave tape and two- and three- dimensional woven or braided structural shapes offer the potential for automated equipment construction with indefinite ambient conditions storage. Braiding will be evaluated for building structural elements such as frames, spars and longerons and for fabricating large components such as a wing box.

Designers are beginning to develop design concepts that exploit the potential to reduce cost with new fabrication methods and automated fiber placement processes. Automated fiber placement is a process that provides inprocess consolidation of multiple ribbonized prepreg tows and uses robotic machine technology to control tow dispensing, location and direction. Individual towpreg ribbons have cut and add dispensing control so that each tow in a 32 tow band can be applied or removed without stopping the delivery head as it applies material. Automated fiber placement machines can be used to fabricate integrated structures by rapidly placing towpreg on complex shape tools or mandrels. Stiffeners can be wound

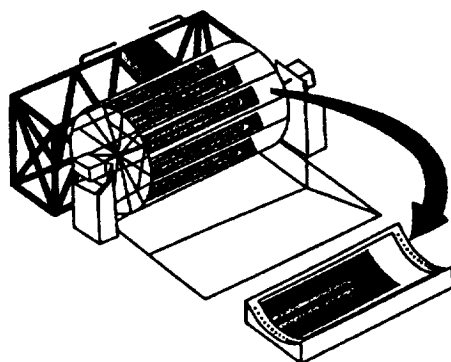


into grooves in a mandrel that are oriented in any suitable stiffener orientation and then the skin is wound over the stiffeners to produce a stiffened shell structure with minimum part count, tool requirements, and a single cure cycle for the complete part. An ACT contract with Hercules Inc. (19) will use their automated fiber placement equipment shown in Figure 11 to demonstrate the cost effectiveness of the process to



**Figure 11. - Multi-axis fiber placement machine.**

fabricate multiple panels similar to the Boeing fuselage keel quadrant panels shown in Figure 12. Advantages of the fiber placement process are provided in Table 3.



- Minimize splices and part count
- Automated batch processing
- Efficient ply tailoring
- Low-cost material form
- Compatible with all design families

**Example: Keel quadrant skins**

**Figure 12. - Process efficiency of quadrant approach**

- 
- Ply thickness control
    - Thin plies, prepreg tape equivalent
    - Constant thickness
  - In-process compaction
  - Controlled fiber angles
    - Nongeodesic paths
  - Potential low cost material form  
(In-line prepreg tow)
  - 360° part fabrication
    - Longitudinal joint elimination
    - Fastener reduction
- 

**Table 3. - What fiber placement offers.**

## Summary

NASA, DARPA (20), DoD and industry all are seeking lower acquisition cost approaches to applications that will extensively increase the volume usage of composites in the next decade. Trade studies in Phase A of the ACT program have indicated that several composite manufacturing approaches will provide substantial weight savings at equal cost to aluminum for large primary structure. As composites processes are optimized and low cost textile material forms are introduced to the market place, composites will be less expensive than equivalent metallic structures. Cost effective composites will broaden opportunities for introduction of fatigue and environmental degradation resistant materials that exhibit near zero thermal expansion coefficient to broad sectors of industrial, commercial, automotive and sporting goods applications.

## References

1. Bohon, H. L. and Davis, J. G., Jr.: Composites for Large Transports-Facing the Challenge Aerospace America, June 1984.
2. Dow, M. B.: The ACEE Program and Basic Composites Research at LARC (1975 to 1986). Summary and Bibliography. NASA RP 1177, October 1987.
3. Coggeshall, R. L.: Boeing/NASA Composite Components Flight Service Evaluation. NASA CR-181898, 1989.
4. Anon: Advanced Organic Composite Material for Aircraft Structures-Future Programs. ASEP/NRC Report. National Academy Press, 1987.
5. Wilson, R. D.: Advanced Composite Development For Large Transport Aircraft. Presented at the 16th Congress of the International Council of Aeronautical Sciences (ICAS), Jerusalem, Israel, Aug. 1988.
6. Povinelli, F. P.; Klineburg, J. M.; and Kramer, J. J.: Improving Aircraft Energy Efficiency. *Astronautics & Aeronautics*, vol. 14, no. 2, pp. 18-31, Feb. 1976.
7. Dexter, H. B.: Long-Term Environmental Effects and Flight Service Evaluation of Composite Materials. NASA TM 89067, January 1987.
8. The Boeing Commercial Airplane Company: High-Speed Civil Transport Study. NASA CR-4234, Sept. 1989.
9. Schlack, M.: Composite manufacturing takes a giant leap forward. *Plastics World*, Jan. 1990, pp 46-48.
10. Henderson, B. V.: Boeing's B-2 Facility Produces Largest Composite Structural Parts Ever Made. *Aviation Week and Space Technology*, Sept. 17, 1990, pp 59-62.
11. Bohon, H. L. and Davis, J. G.: Advanced Composites Technology-Status and Opportunities. FIBER-TEX 89, Greenville, SC, Oct. 30-Nov. 1, 1989.
12. Dexter, H. B. and Funk, J. G.: Impact Resistance and Interlaminar Fracture Toughness of Through-the-Thickness Reinforced Graphite/Epoxy. AIAA Paper No. 86-1020, April 1990.
13. Dexter, H. B. and Maiden, J.: Application of Textile Material Forms to Composite Aircraft Structures. FIBER-TEX 87, Greenville, SC, Nov. 3-6, 1987. NASA CP-3001, April 1988.
14. Smith, D. L. and Dexter, H. B.: Woven Fabric Composites With Improved Fracture Toughness and Damage Tolerance. FIBER-TEX 88, Greenville, SC, Sept. 13-15, 1988. NASA CP-3038, June 1989.
15. Dow, M. B. and Smith, D. L.: Damage-Tolerant Composite Materials Produced by Stitching Carbon Fabrics. International SAMPE Technical Conference, Vol. 21, pp. 595-605, 1989.
16. Palmer, R. and Cursio, F.: Cost-Effective Composites Using Multi-Needle Stitching and RTM/VIM. FIBER-TEX 88, Greenville, SC, Sept. 13-15, 1988. NASA CP-3038, June 1989.
17. Baucom, R. M. and Marchello, J. M.: LaRC Powder Prepreg Towpreg System. International SAMPE Technical Conference, Vol. 35, pp. 175-188, 1990.
18. Baucom, R. M. and Marchello, J. M.: LaRC Dry Powder Towpreg System. NASA TM 102648, March 1990.
19. Anderson, R. L. and Grant, C. G.: Advanced Fiber Placement of Composite Fuselage Structure. ACT Conference, Seattle, Washington, Oct. 29-Nov. 1, 1990.
20. Leonard, L.: Composites at Sea. *Advanced Composites*, March/April 1990, pp 38-58.

## RAPID INDUCTION BONDING OF COMPOSITES, PLASTICS, AND METALS

John D. Buckley  
Langley Research Center  
Hampton, VA

Robert L. Fox  
Langley Research Center  
Hampton, VA

## ABSTRACT

The Toroid Bonding Gun is an induction heating device. It is a self contained, portable, low powered (0.1-2 kilowatt) induction welding system developed at Langley Research Center for bonding or joining plastic, ceramic, or metallic parts. Structures can be bonded in a factory or in the field. Bonding times for laboratory specimens have been cut by a factor of 10 to 100 times compared to standard press or autoclave bonding. This type of equipment allows for applying heat directly to the bond lines and/or the adhesives without heating the entire structure, supports, and fixtures of a bonding assembly. The induction heating gun originally developed for use in the fabrication of space structures has been expanded. Gangs of bonders are now used to rapidly join composite sheet and structural components. Depending on the type of bonding equipment, composite structures can be joined in a factory or in the field. Other NASA-developed applications of this bonding technique include the joining of thermoplastic composites, thermosetting composites, metals, and combinations of these materials. Inductron Corporation has been granted an exclusive license to practice the technology. Inductron has developed techniques and tools for annealing and brazing metals, manufacturing thermocouples, shrinking Cryo-fit (TM) hydraulic couplings, and flexible induction heads for complex structure repair/ manufacture. Inductron will have a booth to display commercially developed hardware at Tech 2000.

## INTRODUCTION

In the fabrication of large space structures (Figure 1), a portable induction heater requiring minimal or no external cooling is desired when sealing seams and joining components. This portable heater used little power and is controlled by miniaturized components. The toroid joining gun is a self-contained, portable low powered (.1-2 Kilowatts) welding system developed at Langley Research Center for joining plastics, metal, and ceramic parts using induction heating techniques (Figure 2). The device developed for use in the fabrication of large space structures (Figure 3) can be used in any atmosphere or in a vacuum. Components can be joined in situ, whether on earth or on a space platform. The expanded application of this welding gun is in the joining of thermoplastic composites, thermosetting composites, plastics, metals, and combinations of these materials, annealing and brazing of metals, making of thermocouples, and shrinking of Cryo-fit hydraulic couplings.

Induction heating is well known in industry. When a current passes through a conductive coil, a magnetic flux is produced inside and around the coil. When a mass of conducting material is moved in this magnetic field or is subjected to a changing flux, induced currents, called eddy currents, circulate in the mass. In most instances, eddy currents are, wasted energy, but in induction heating they are actual producers of heat.

The objective of this bonding system was to provide a method and apparatus for joining thermoplastic matrix composites and other compatible materials which could be used in outer space, on earth, or in motionless surroundings. The ability of a low-powered toroid gun (Figure 5) to heat metal, and a recently developed non-metal ceramic, is used to inductively join pieces of thermoplastic and thermoplastic composite. In a toroid pole piece, magnetic flux remains inside the toroid core when the system is energized. To divert the path of the magnetic flux from the toroid to an adjacent susceptor, the toroid must be altered. This alteration is accomplished by cutting a segment out of the toroid and placing the air gap in the toroid on the

surface of a plastic matrix composite sandwich consisting of a susceptor positioned between the two composite components to be joined (Figure 5). When the toroid is energized, flux will flow through the toroid, through the plastic composite (which is transparent to magnetic flux) into the susceptor back through the plastic composite into the toroid. Alternating current produces inductive heating instantly in the susceptor causing the plastic interfacing on either side of the susceptor to melt and flow into perforations made in the susceptor forming the joint. Joining is accomplished in minutes.

Inductron Corporation has expanded the applications of the Toroid Bonding Gun to include brazing and heat treatment of metals, shrinking of Cryo-fit hydraulic couplings, making of thermocouples, and repair/fabrication of ceramic, metallic, and plastic components. To accomplish the expanded applications of the U.S. Government patent, Inductron has modified and improved circuit and head design, included temperature control, temperature monitoring, timers, and in some cases, included temperature recording.

## BONDING AND TESTING

### Shear Specimen Bonding

Overlap shear specimens were bonded in a configuration conforming to the American Society for Testing Materials (ASTM) standards D1002 and D3163. The technique similar to that used for spot welding metallic structures was used for rapid bonding of lap shear specimens made of thermoplastic composites, thermoset composites, metals, and combinations of these materials (Figure 4).

The rapid bonding equipment for laboratory shear specimens is shown in Figure 6. The press is identical to that for conventional specimen bonding, as are the load cell and temperature and load indicators. Replacing the conventional heated platens is a toroidal high frequency induction heater and its power controller. The specimen is located in a fixture for ease of alignment. When power is applied, the toroid rapidly heats the susceptor, such as a metal screen, which has been impregnated with a thermoplastic adhesive or is sandwiched between thermosetting adhesive films. For lap-shear specimens, the metallic or fiber-reinforced plastic composite material adherends are placed above and below the susceptor in the sample fixture, and bonding pressure is applied. The susceptor heats the adhesive or thermoplastic composite adherend rapidly, usually within a minute, to the bonding temperature. The heat is maintained from one to several minutes to promote adherend joining. When power is turned off, the specimen rapidly cools to a temperature below which the adhesive or thermoplastic composite is sufficiently set, and pressure is removed. Some of the composite materials tested are shown in Table 1. When rapid bonding was compared directly with laboratory-press/heated-platen bonding the strengths of the toroid bonds were found to be significantly higher than those obtained by press bonding (Table 2). This process is more controllable and more energy conserving than conventional bonding with heated platens or an autoclave (1, 2).

## BONDING APPLICATIONS-NASA

In addition to the lap shear mechanical property tests obtained on adhesives, a number of applications have been found for using this technology.

### Panel Bonding

An overall view of the Rapid Adhesive Bonding (RAB) equipment of overlap panel assemblies is shown in Figure 7. Six air-cooled toroid induction heaters are mounted with a load cell on a support structure, which in turn is mounted to a transverse alignment slide. That slide is located on a longitudinal traversing carriage, which travels along a standard machine bed. The power supply controller and load indications are on the transverse slide, while thermocouple sensor monitors of panel specimen temperature on the bed. The panels to be bonded, 2 feet long in this figure, are located in a simple insulated panel support. The power supply shown (which uses approximately 1000 watts or less at 60 Hz and 120 volts for each toroid) provides the required circuitry for a self-tuning induction-heating circuit which operates at 30,000 to 100,000 Hz. Susceptors concentrate the induction heating effect in the bond line. The most successful of these developed

for this application are the flattened 1008 carbon steel screen and a perforated 410 stainless steel foil.

**Temperature** - Temperature monitoring in the adhesive bond line is considered to be an important requirement of this Induction Bonding Process, because heating is concentrated in the bond line in all applications in which a susceptor is used. A series of thermocouples are positioned an inch apart in the bond line of a test specimen for each of the materials to be bonded.

**Bonding Procedure** - The bonding procedure consists of installing the parts to be bonded in a fixture, positioning the toroid heads over a dummy "on-ramp" specimen of the same material and geometry as the work, bringing the dummy specimen (with a thermocouple in the bond line) to the bonding temperature, and initiating a traversing carriage movement at the required rate.

Graphite/epoxy panels bonded with the HT-424 adhesive are shown in Figure 8. They were made with and without susceptors in the bond line. An important finding in this application of RAB is that Graphite/epoxy laminates can be heated directly in the induction field of the toroid heaters without a susceptor. Some lap shear specimens, cut from a graphite/epoxy panel bonded with HT-424 epoxy/phenolic adhesive and with the steel screen susceptor, were tested at room temperature and at 180°F (in accordance with ASTM D1002 and D3163). Other samples were tested at RT and 180°F, after 1000 thermal cycles from -100° and 180°F, at RT and 180°F after a 72 hour boiling water exposure. The RAB process had no degrading effect on shear strength of Gr/Ep/Ht-424/Gr/Ep bonds, compared to standard bonding, and thermal cycling did not significantly degrade these properties. The water boil exposure degraded bond strengths about 35 percent at room temperature and 28 percent at 180°F. This degradation is approximately the same as that noted for the HT-424 adhesive in the supplier's literature.

#### Rapid Strain Gauge Attachment

A new method has been developed for bonding film gauges (e.g., strain gauges) to high reluctance, low reluctance, and no reluctance structures.

Conventional methods for bonding film gauges to be operated at elevated temperatures are typically time consuming due to clamping and baking procedures. Because it is impractical to utilize an oven for curing purposes in the field, different procedures and equipment must be implemented.

Previous methods of mounting gauges require the use of contact adhesives and epoxies. Contact adhesives for mounting gauges are limited to low humidities during installation, low temperature during use, and 90 days of useful data measurements. Epoxies require the use of heat lamps, hair dryers, and electric heater devices to generate heat to cure the epoxies. The use of epoxy as a strain gauge adherent requires substantial time and power.

A new method of bonding strain gauges that reduces adhesive cure schedules from a matter of hours to minutes has been developed. This procedure, which incorporates the use of a toroid joining gun, is primarily applicable to elevated (up to 204°C) temperature strain gauge requirements. The toroid gun can concentrate high temperatures in localized areas through the induction heating principles. When bonding a sensor to the surface of the insulator, such as fiberglass, which has no reluctance and poor thermal conductivity, a ferric material in the form of a thin (1/8") plate is placed between the sensor and the energy source.

The ferric plate generates inductive heat which is transferred to the gauge region to activate the adhesive. The heat is obtained from the induction bonder which is placed on the gauge with approximately 5 pounds of pressure per square inch and energized for 2 minutes (Figure 10). Pressure is maintained for an additional 1 minute during cool down. For this process, a hand held gun is used, as shown in Figure 9C. Once the gauge area has cooled down, leads are attached, and testing begins.

Test data shows that strain gauges bonded to structural specimens using the toroid induction bonder produce results equal to the data obtained using standard bonding techniques (4). The advantages of using

the toroid joining gun over existing methods are lower power, rapid bonding, lightweight equipment, localized heat on the strain gauge adhesive, and portability.

### One-Step Joining and Electric Conduction

This process relates to electromagnetic heating and, in particular, to a fastener to be used between two thermoplastic or composite objects for induction heating and joining purposes.

Induction heating using a metal fastener is a common method of joining. The metal fastener may be a solid sheet of eddy current conducting metal placed adjacent to the materials to be joined, or an apertured or screened sheet (metallic or nonmetallic) placed between the materials. Electromagnetic induction heats the metal when one of several known induction heating methods is used; the heat generated softens and joins the nonmetallic material.

Often the step following the joining of thermoplastics is that of providing electronic circuitry thereupon. An electrical conduction system is usually separately fabricated and fastened to the joined materials. No means have been provided for inductively heating sheets and simultaneously fixing electronic circuitry into the system.

In the one-step dual joining technique (5), a screen is placed between a thermoplastics at the area to be joined and is inductively heated using the toroid joining gun. The materials to be joined soften and join at the apertures of the screen. After joining, the screen lends structural support, and the copper wires therein may be used to conduct electricity (Figure 11a and b). Insulation allows conduction without the danger of short circuiting, and the thermoplastics, themselves insulators, ensure safe conduction. The conducting wires may lead to any electronic component, and render separate electronic circuitry unnecessary.

Typically a 26-mesh screen of carbon steel inductor wires is integrated with two or more insulated copper wires. This integrated screen is placed between two sheets of plastic. The toroid joining gun produces sufficient heat to join the sheets of plastic. After cooling, the copper wires are able to conduct electricity through the now joined sheets of plastic.

Conducting wire or bundles of wires may be used. Some inductor wires are necessary for formation of the bond; their number relative to conductor wires is chosen according to the desired heat of joining and conductivity of the joined sheets. Since the sheets bond to each other around the wires, the number of conductor wires present within the sheets does not affect the strength of the final bond.

A large fabrication of sheets and wire screen may serve as a wall panel in, for example, a house (Figure 11a and b). The copper wires lead to conventional electrical outlets or switches.

In an even larger fabrication, copper wires within sheets making up cylindrical tubes connecting components of a large space structure, could transmit data in electrical form between components.

## **BONDING APPLICATIONS - INDUCTRON**

### Rapid Field Repair with Susceptors

Inductron has developed a device (Flexible Heating Head) to meet a need for a better way of repairing aircraft windscreens, metallic and composite structures (Figure 12a, b, and c). Conventionally, windscreens and composite panels are repaired by using mechanical fasteners to clamp a patch, by fusing the patch with the damaged structure (e.g. windscreens), or by adhesive bonding of the patch. Each method has drawbacks; mechanical fasteners require hole preparation and special hardware, fusing thermoplastic composite material requires fixturing and time for curing.

The simply operated flexible induction welding gun employs magnetic flux to a susceptor generating heat

to the parts to be joined. In a typical aircraft repair, the only preparation required (e.g., composite, metal,...) is cutting a susceptor and patch slightly larger than the damaged area. The susceptor is placed between the damaged component and the patch. The susceptor is impregnated with a thermoset or thermoplastic adhesive. The flexible welding gun is positioned directly above the damaged area. When the flexible head is energized, the alternating current produces inductive heating in the susceptor, the susceptor transmits the heat to the bond line, causing the adhesive to melt, flow, and cure (thermoset) forming a bond (Figure 13).

#### Rapid Field Repair Without Bond Line Susceptors

Inductron has developed a repair process for damaged ceramics, metals, and plastics eliminating the need of bondline susceptors (Prior Art). This is accomplished through the use of a flexible ceramic susceptor. The susceptor is placed above the damaged area, which is comprised of a patch, adhesive and damaged component. An induction welding gun (flexible head) is positioned directly above the susceptor. When the induction welding gun is energized, the alternating current produces inductive heating in the susceptor, the susceptor transmits the heat to the patch, adhesive and damaged component causing the adhesive to melt, flow, and cure (thermoset) forming a bond.

### **COMMERCIAL EQUIPMENT - INDUCTRON**

#### Making of Thermocouples

Inductron Corporation has developed a rapid, reliable, inexpensive method of manufacturing thermocouples. A thermocouple is composed from two wires of dissimilar metals joined at one end.

Inductron has developed a brazing process to join the dissimilar metals of a thermocouple. Inductron Corporation has redesigned the Toroid Bonding Gun to include a susceptor in the gap. Once the bonding head is energized alternating current flows in the susceptor, a braze alloy and flux are now used to tin the susceptor. The two thermocouple wires are twisted and placed in the soft brazing material, on the susceptor, for 5 seconds and removed forming a thermocouple (Figure 14).

#### Flexible Susceptor

Inductron Corporation has developed a repair process for metals, ceramics, and composites utilizing a flexible ceramic susceptor.

The flexible susceptor can be inductively heated outside the bondline as a means for joining metals, plastics, and ceramics using thermoset/thermoplastic adhesives. The flexible susceptor may be used as a heating element (susceptor) in the bondline to melt thermoplastics or for curing thermoset adhesives.

### **TOROBRAZER**

A new concept (TOROBRAZER) has been developed by Inductron Corporation for brazing and annealing bandsaw blades utilizing an induction heating process.

The TOROBRAZER can concentrate high temperature levels in localized areas through induction heating principles. A half inch gap has been cut from the toroid. When the power source is energized, a current passes through the coil generating a magnetic flux which travels to the gap of the toroid. The flux is then conveyed to the saw blade through the chassis top plate which is transparent to magnetic flux. Resulting eddy currents produced in the blade generate heat sufficient for brazing (Figure 16) and annealing the blade joint.

#### Coupling Head

Inductron Corporation has developed a coupling head (Figure 15a) for heating metallic heat-to-shrink couplings and fittings (Figure 15b).

The coupling head is capable of reliably generating high levels of heat within the coupling using an induction heating process. A gap has been cut in the toroidal core. When the power source is energized a current passes through the core to the toroidal gap. The resulting hysteresis losses produced in the coupling (positional in the gap) generate heat sufficient for shrinking the coupling (440°F). The power source is then automatically deenergized by the temperature controller.

## CONCLUSION

Induction Bonding concepts and equipment have been developed at NASA and Inductron. Bonding times for standard ASTM overlap shear specimens can be cut by a factor of 10 to 100 compared to standard press or autoclave bonding. High lap shear strengths can be generated with a range of adherend materials (including metals and polymer matrix composites) and adhesives (both thermoplastic and thermosetting). Short term thermal cycling and water boil exposures have shown environmental stability for these rapid bonds, including those which contain steel screen or stainless steel foil susceptors in the bond lines.

The Induction Bonding concepts were extended to continuous seam bonding of metallic and composite panels with promising results for bonding of both like and unlike adherends. Rapid bonding of other geometries such as face sheets of fiber-reinforced polymer matrix composites or titanium alloy to titanium honeycomb were proven feasible.

The inherent portability of Induction Bonding equipment suggested that field repair procedures for adhesive bonding of damaged metallic, polymeric, or composite structures are possible. Initial development of these procedures showed that field bonding of patches of titanium alloy and graphite/epoxy composite materials could be bonded to typical aircraft panels.

Furthermore, variations of the Induction Bonding process can be used to repair polycarbonate or acrylic windscreen material and hydraulic tubing; anneal metals; fabricate thermocouples; mount strain gauges; and braze metals.

## REFERENCES

1. Buckley, J.D.; Fox, R.L.; and Swaim, R.J.: Plastic Welder, NASA Tech Brief LAR-12540, Summer 1980.
2. Stein, B.A.; Tyeryar, J.R.; Fox, R.L.; Sterling, E.S., Jr.; and Buckley, J.D.: Rapid Adhesive Bonding for Metals and Composites, NASA Tech Brief LAR 13066, Spring 1984.
3. Spinoff 1984: Portable Welder, p. 98; NASA Office of External Relations, Technology Utilization and Industry Affairs Division by Haggerty, J.J., July 1984.
4. Schott, T.D.; Fox, R.L.; and Buckley, J.D.: Inductive Energy for Rapid Strain Gauge Attachment, U.S. Patent 4,767,484, August 1988.
5. Buckley, J.D.; Swaim, R.J.; and Fox, R.L.: One-step Dual Purpose Joining Technique, U.S. Patent 4,313,777, February 1982.



TABLE I

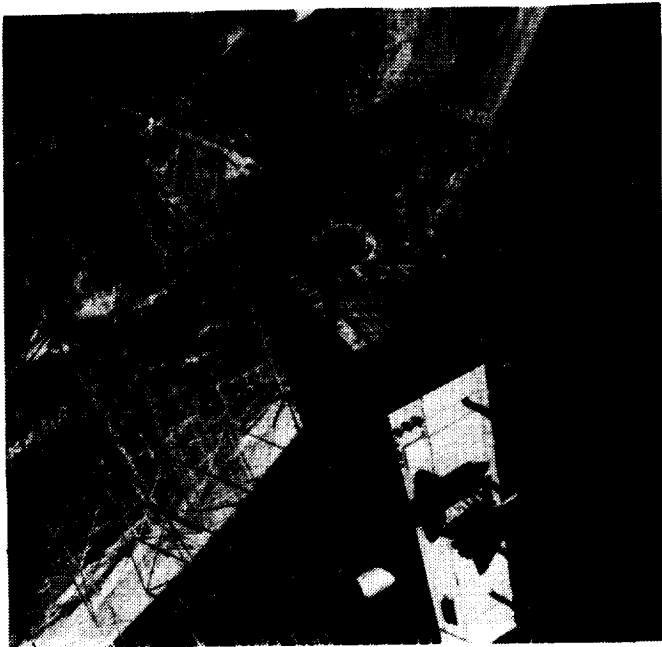
**COMPARISON OF BONDING PARAMETERS  
OF RAPID ADHESIVE INDUCTION BONDING  
WITH CONVENTIONAL BONDING**

Composite Material	Bonding Parameters							
	Press., psi		Temp., °F		Heat Rate, °F/min		Hold Time, min	
	Conv.	RAB	Conv.	RAB	Conv.	RAB	Conv.	RAB
Vinyl Ester/Gp	50	10	350	350	9	350	15	2
Epoxy/Gp	50	10	350	350	9	350	15	2
PEEK/Gp	200	10	750	750	9	300	15	2
PEI/Gp	100	50	625	625	9	1200	15	2
LARC-TPV/Gp	100	50	650	650	9	500	15	2

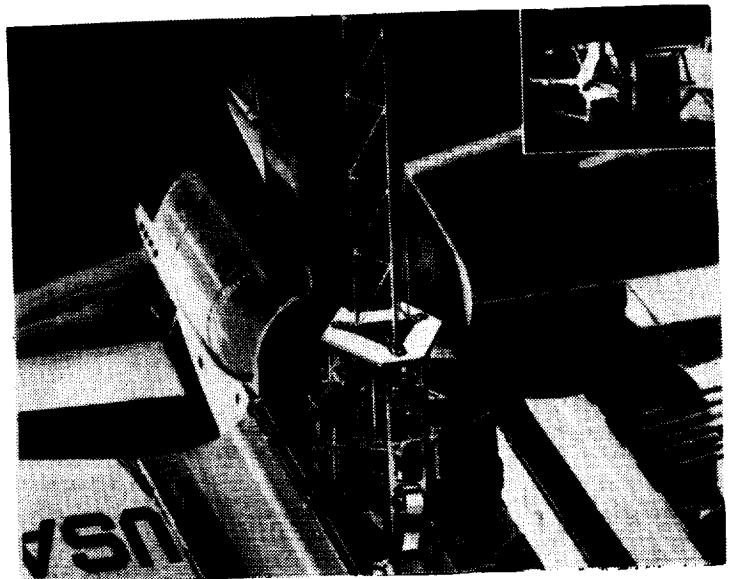
TABLE II

**COMPARISON OF RAPID ADHESIVE INDUCTION  
BONDING WITH CONVENTIONAL BONDING**

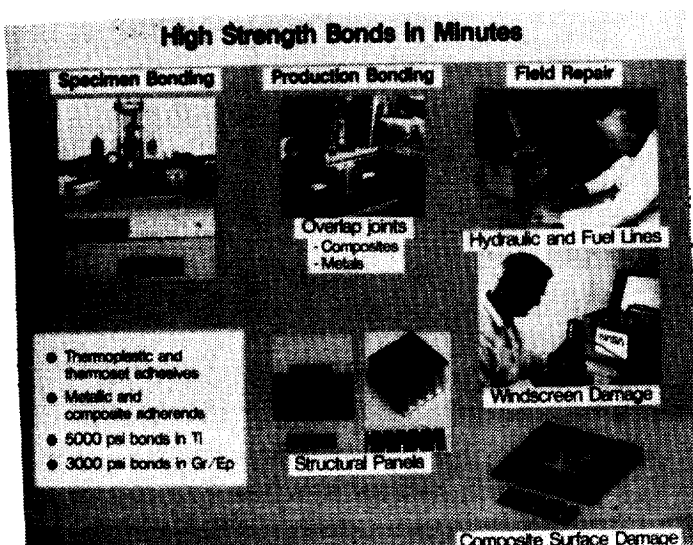
Composite Material	Overlap Shear Strength at Room Temp., psi	
	Conventional Bonding	Toroid Induction Bonding
Vinyl Ester/Gp	3,000	3,800
Epoxy/Gp	3,000	4,500
PEEK/Gp	4,500	7,000
PEI/Gp	4,000	5,100
LARC-TPV/Gp	4,780	4,850



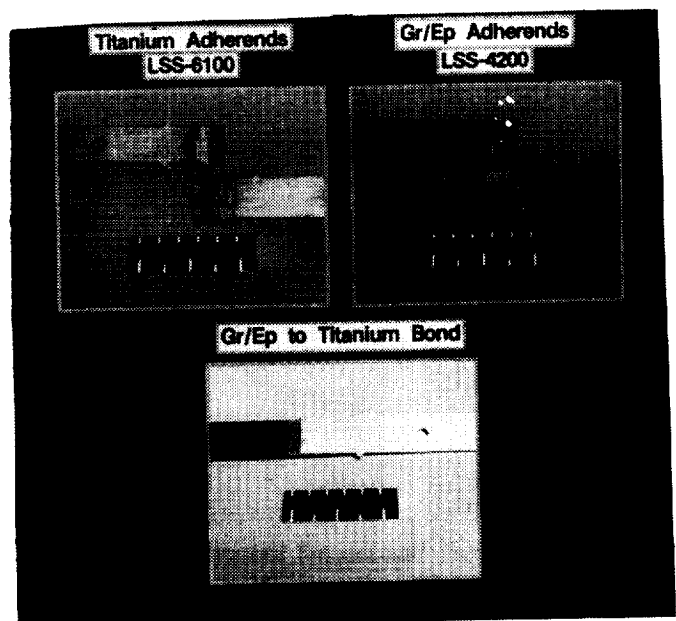
**Figure 1.** Concept-Space Shuttle participating in construction of large area space structure



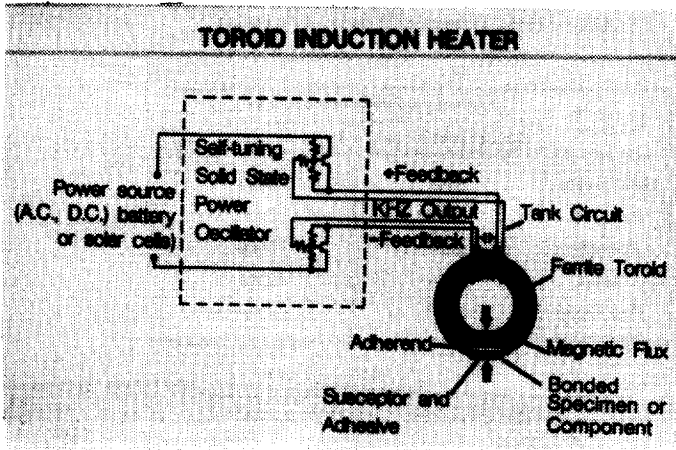
**Figure 3.** Induction bonding system used in space manufacturing of beam structures



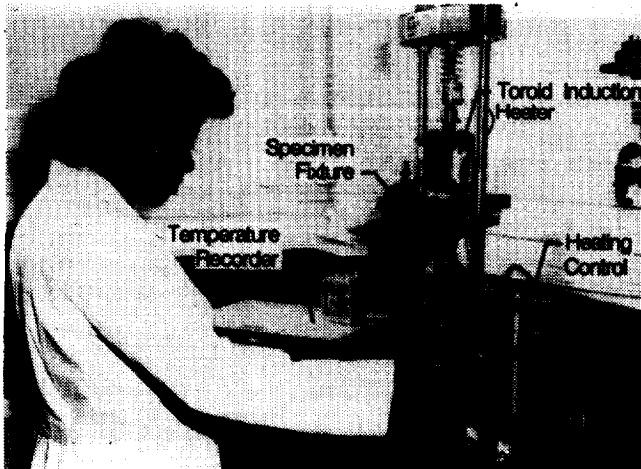
**Figure 2.** Induction heating for high strength bonds in minutes



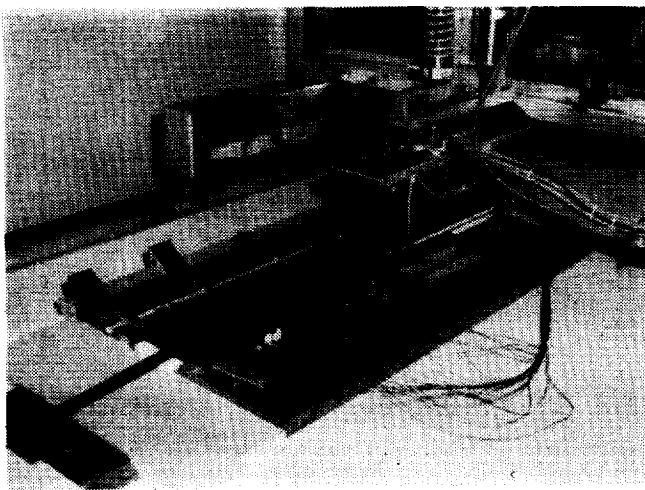
**Figure 4.** Typical induction bonded overlap shear test specimens



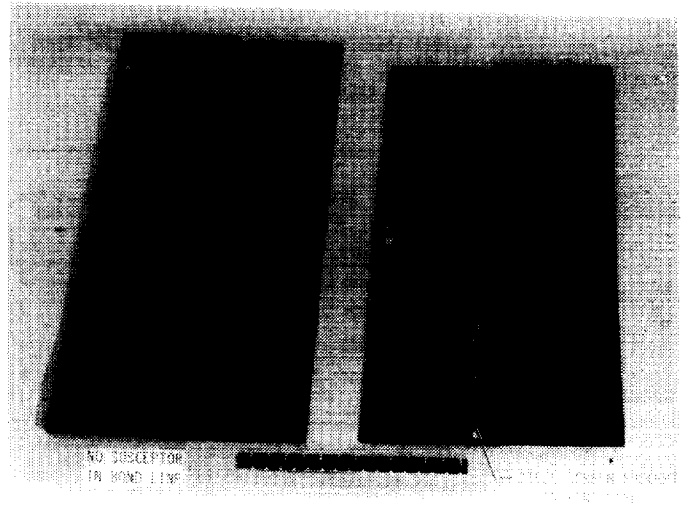
**Figure 5.** Schematic of toroid induction heating concept



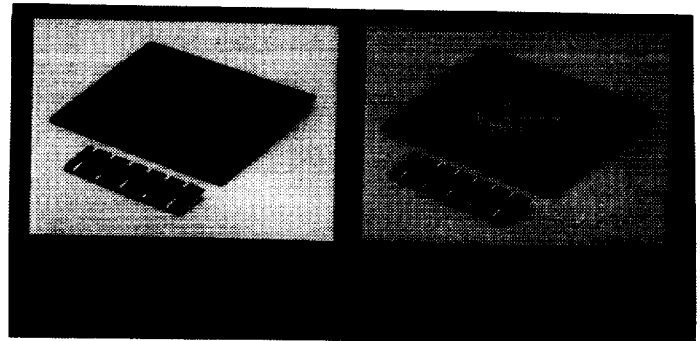
**Figure 6.** Toroid induction heater specimen bonding equipment



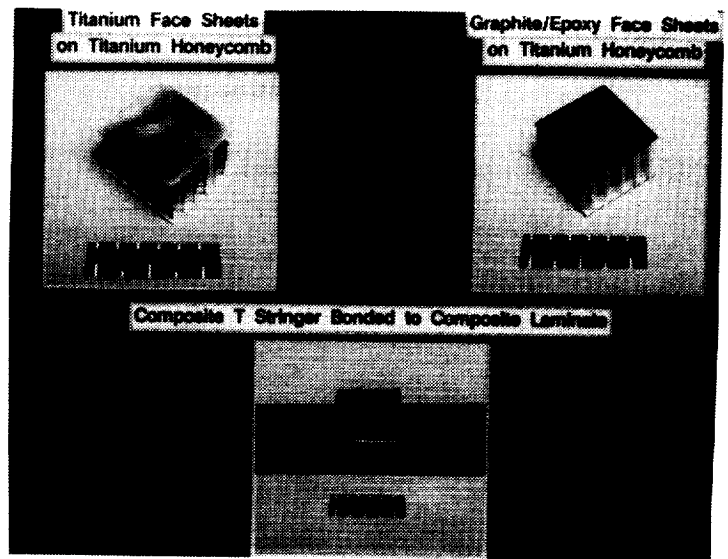
**Figure 7.** Six-headed induction seam bonding machine



**Figure 8.** Seam bonded graphite epoxy panels (23 in. x 10 in. x .255 in.)  
1. No susceptor 2. screen susceptor



**Figure 9(a).** Repair of graphite/epoxy composite laminate



**Figure 9(b).** Manufacture of honeycomb and T stringer structures

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



Figure 9(c). Repair of the canopy of an F-84 aircraft

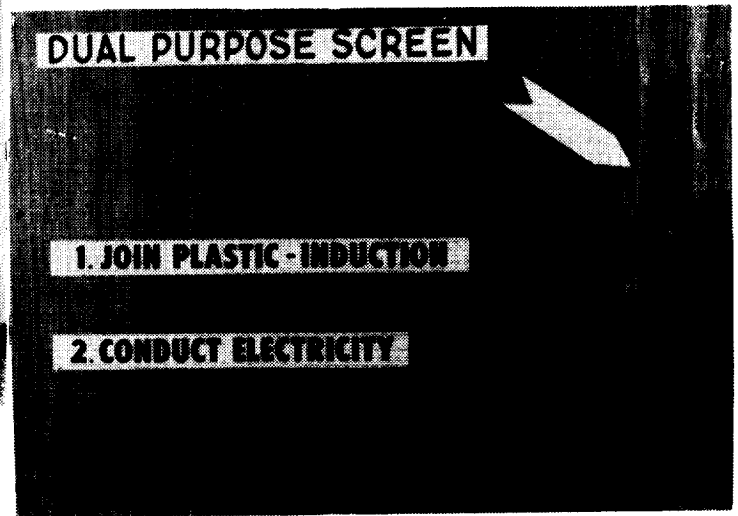


Figure 11(b). Dual purpose bonding technique for bonding structures and conducting electricity

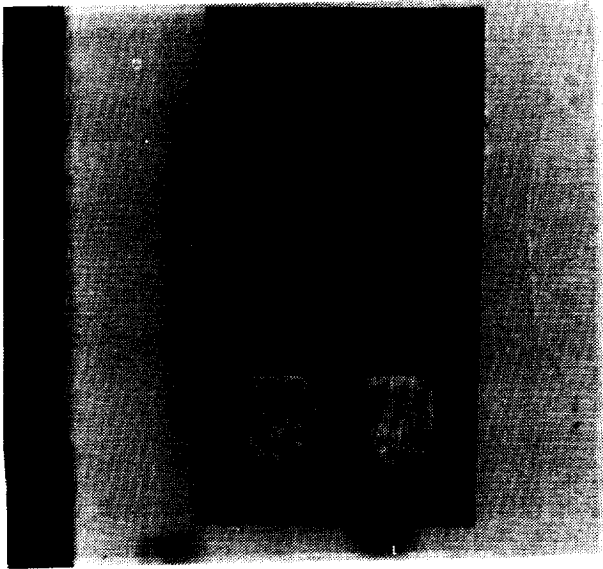


Figure 10. Induction bonded strain gauge on fiberglass composite test specimen

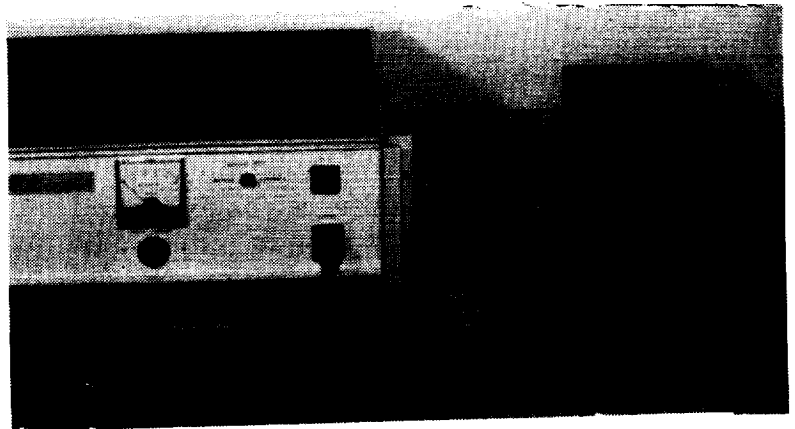
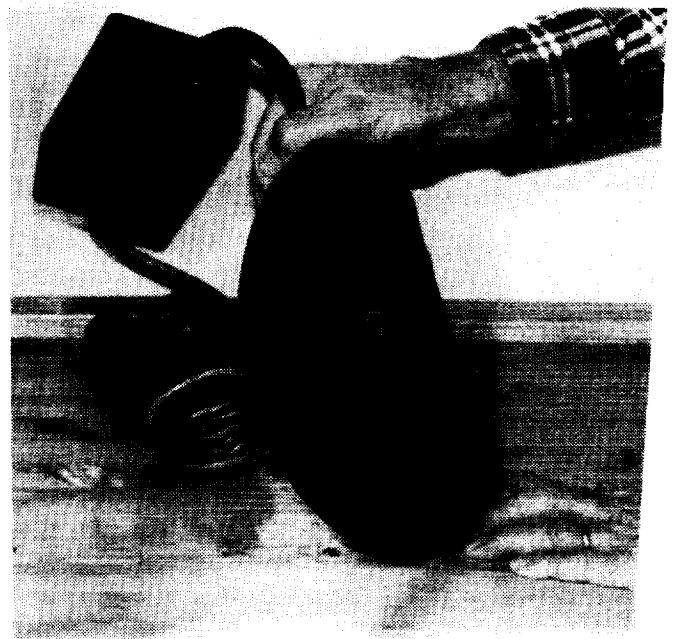


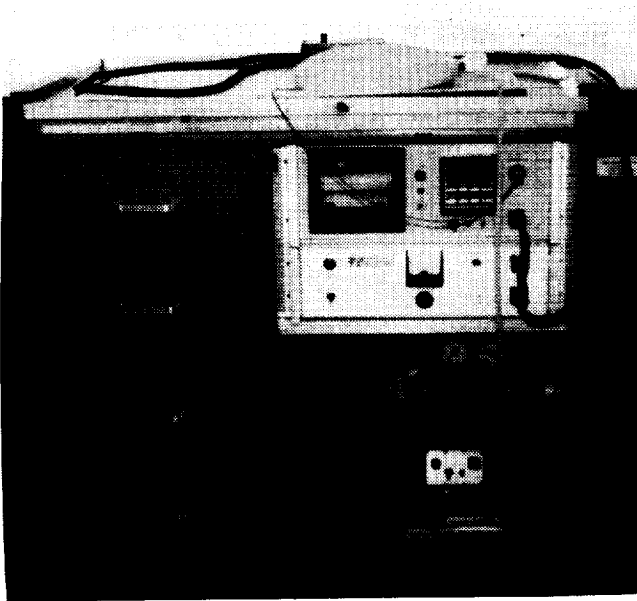
Figure 12(a). Flexible Head and 2KW Power Supply



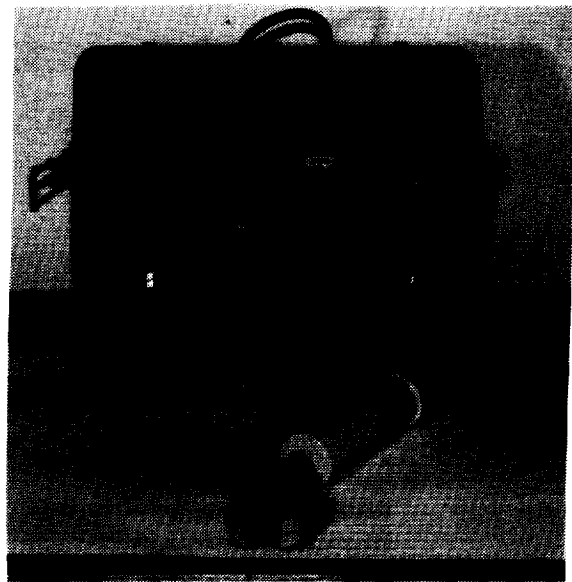
Figure 11(a). Dual purpose bonding technique in a wall section



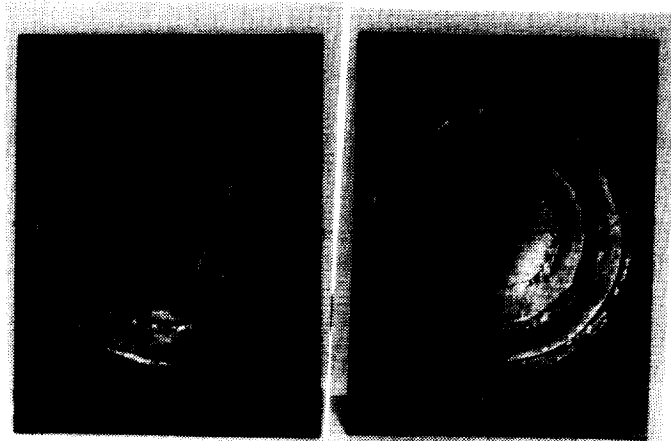
239 Figure 12(b). 12" Flexible Head



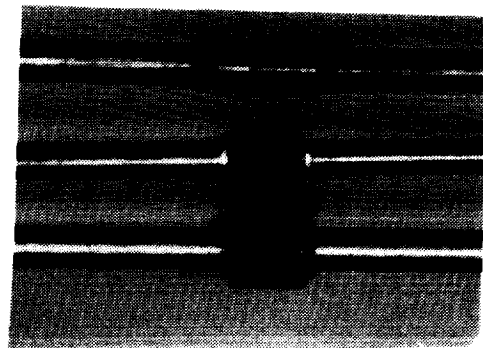
**Figure 12(c). Field Repair System**



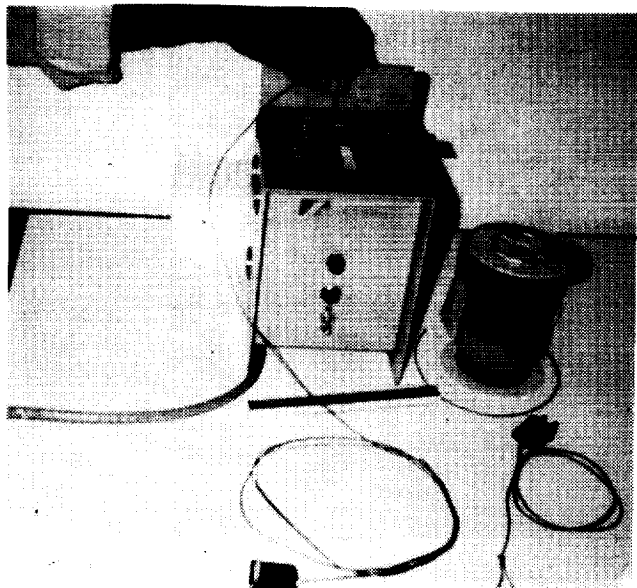
**Figure 15(a). Heat to Shrink Coupling Head**



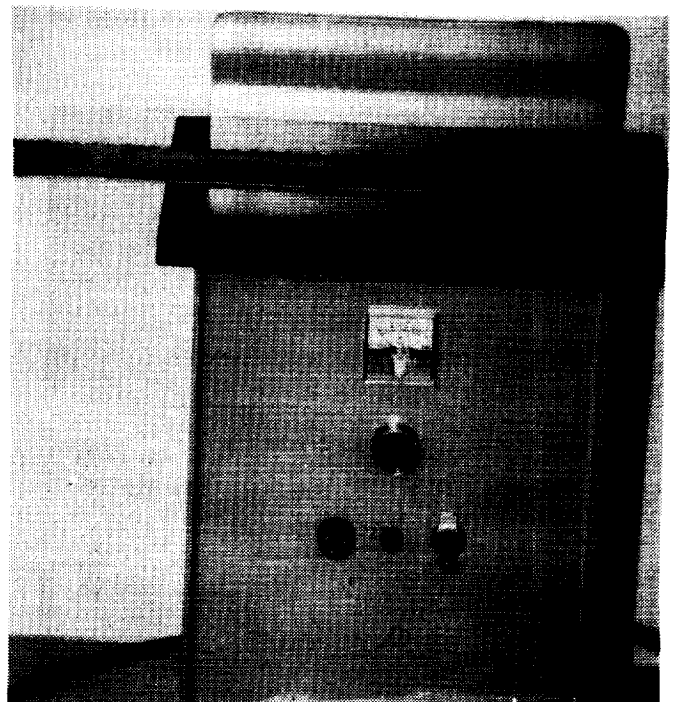
**Figure 13. Composite Repair**



**Figure 15(b). Heat to Shrink Coupling**



**Figure 14. Making of Thermocouples**



**Figure 16. Induction Anneller**

## RELIABILITY AND RISK ASSESSMENT OF STRUCTURES

C. C. Chamis  
NASA Lewis Research Center  
Cleveland, OH 44135

## ABSTRACT

Development of reliability and risk assessment of structural components and structures is a major activity at Lewis Research Center. It consists of five program elements: (1) probabilistic loads, (2) probabilistic finite element analysis, and (3) probabilistic material behavior, (4) assessment of reliability and risk, and (5) probabilistic structural performance evaluation. Recent progress includes: (1) the evaluation of the various uncertainties in terms of cumulative distribution functions for various structural response variables based on known or assumed uncertainties in primitive structural variables, (2) evaluation of the failure probability, (3) reliability and risk-cost assessment, and (4) an outline of an emerging approach for eventual certification of man-rated structures by computational methods. Collectively, the results demonstrate that the structural durability/reliability of man-rated structural components and structures can be effectively evaluated by using formal probabilistic methods.

## INTRODUCTION

It is becoming increasingly evident that deterministic structural analysis methods will not be sufficient to properly design critical components in future structures in general and aerospace structures in particular. These structural components are subjected to a variety of complex, and severe cyclic loading conditions, including high temperatures and high temperature gradients. Most of these are quantifiable only as best engineering estimates. These complex loading conditions subject the material to complex and coupled nonlinear behavior which depends on stress, temperature, and time. Complex and coupled nonlinear material behavior is nonuniform, is very difficult to determine experimentally, and perhaps impossible to describe deterministically. In addition, hot rotating structural components for aerospace propulsion (engines) are relatively small. Fabrication tolerances on these components, which in essence are small thickness variations, can have significant effects on the component structural response. Fabrication tolerances by their very nature are statistical. Furthermore, the attachment of components in the vehicles integrated structural system generally differs by some unknown amount from that assumed for designing the component. In summary, the fundamental aspects -- (1) loading conditions, (2) materials behavior, (3) geometric configuration, and (4) supports (attachments) to integrated structures inherently include a variety of uncertainties of unknown magnitude.

There are generally two approaches to handle this wide variety of uncertainties: (1) current practice, and (2) probabilistic evaluation. Current practice is adequate where the new engine structure is not very different from an existing one. However, this approach is costly and requires long time-schedules for future structures which will be entirely different from any existing ones. The second approach is to develop probabilistic structural analysis methods where the uncertainties in all the parameters of the four fundamental aspects are described by appropriate probability functions.

Development of the probabilistic structural analysis methodology (PSAM) is an on-going activity at NASA Lewis and is a joint program of in-house and sponsored research (ref. 1). Theoretical considerations, computer codes, and other relevant applications are described in papers presented in conferences (refs. 2 to 6). Activities and progress up to June 1989 are

summarized in reference 7. The objectives of this brief paper are (1) to summarize the fundamental aspects of PSAM and (2) to demonstrate the application of this methodology to a specific example (the reliability/risk of turbine blade components of rocket propulsion systems). The specific example includes the four fundamental aspects (key elements) required in probabilistic structural analysis of future structures, namely: (1) probabilistic loads, (2) probabilistic finite element analysis, (3) probabilistic description of complex coupled nonlinear material behavior, and (4) evaluation of reliability and risk. Throughout the specific example discussion, appropriate comments are included to illustrate the generality of the method and its application to aerospace and other structures in general.

## FUNDAMENTAL CONSIDERATION

Central to the probabilistic structural analysis is the fundamental consideration that: Uncertainties observed in the structural performance (displacements, frequencies, buckling, global fracture toughness, stresses/strains) of structures can be quantified in terms of corresponding uncertainties in basic parameters (primitive variables). The primitive variables are those which are used to describe the structure and its respective environment. For example: (1) structural configuration, (2) boundary conditions (attachments), (3) loading conditions, and (4) material thermomechanical nonlinear behavior.

The uncertainties in these primitive variables are then integrated through structural mechanics to quantify the uncertainties in the global structural responses (displacements) and are decomposed to quantify the uncertainties in local responses (stresses/strains). The concept is schematically illustrated in figure 1. The structural component is the blade which is modeled for finite element analysis. The input uncertainties are the blade loads (centrifugal, pressure, and temperature), geometry and material variables. The output is quantification of uncertainties in structured responses or in local stresses for probable fracture initiation. Brief descriptions are given in subsequent sections on each of these as it is applied to the specific example.

## PROBABILISTIC SIMULATION OF LOADS

The fundamental assumption for the probabilistic simulation of loads is that each individual load condition can be probabilistically synthesized from four primitive parts: (1) steady state, (2) periodic, (3) random, and (4) spike. Each of these parts, except random, is described by a nominal or deterministic portion and a probabilistic perturbation about this nominal portion. The resulting distribution is similar to the schematic, in figure 1 upper left and as further described in reference 7. One justification for synthesizing each loading condition in terms of primitive parts is that experts, over the years, have developed good judgment of the ranges of perturbations about nominal or deterministic conditions. A computer code (Composite Load Spectra) has been developed to synthesize the four parts of each load condition by using (1) available data from various sources of past experience, (2) probability theory, and (3) a dedicated expert system, which includes the information supplied by the experts.

The results from the application of the Composite Load Spectra computer code to probabilistically simulate loads for two blades are summarized in table 1. The comparisons with the measured data are in very good agreement considering the large number of primitive variables (47) required to synthesize these loads. The conclusion is that methods can be developed and are available to probabilistically synthesize complex load conditions for hot aerospace structures and structures in general.

## PROBABILISTIC FINITE ELEMENT STRUCTURAL ANALYSIS

The fundamental assumption for developing probabilistic finite element methods (PFEM) for structural analysis is that the uncertainties in each primitive structural variable can be represented by an assumed probabilistic distribution. Primitive structural variables are those which are used to describe a structure such as: (1) stiffness, (2) strength, (3) thickness and tolerance, (4) spatial location, (5) attachment, and (6) various nonlinear material dependencies (temperature, stress, time, etc) as is schematically illustrated in figure 1 upper right. See also reference 7. Subsequently, the uncertainties in the load conditions (synthesized by the composite load spectra) and the uncertainties in the primitive structural variables are computationally synthesized by performing probabilistic finite element structural analysis to simulate uncertainties in the structural response of a specific structural component or structure. The structural response is generally described in terms of usual quantities such as displacement, frequencies, buckling loads, and structural fracture toughness as was already mentioned.

PFEM has been formalized and integrated into a computer code identified as NESSUS (numerical Evaluation of Stochastic Structures Under Stress). NESSUS contains a library of finite elements and is driven by an expert system. It can be used to probabilistically evaluate all types of structures. Representative results obtained using NESSUS on various structures are given in references 1 to 7 and for the specific example will be discussed in a later section. The combined effects of the primitive variable uncertainties on structural response are generally shown as probability distributions, figure 1 bottom right (stress from NESSUS). The information generated for these probability distributions can also be used to evaluate the sensitivities which influence these distributions. The significant point is the PFEM yields a wealth of information which can be used to evaluate: (1) the uncertainties in the structural response, and (2) the sensitivities which can be used to adjust the design for enhanced probability of success. The important conclusion is that probabilistic finite element methods can be developed and are available to quantify uncertainties in the structural performance of a variety of structures. In addition the sensitivities that influence this performance can be evaluated and ranked.

## PROBABILISTIC SIMULATION OF MATERIAL BEHAVIOR

The fundamental assumptions to probabilistically simulate complex nonlinear material behavior are: (1) a relationship for material behavior can be developed in terms of primitive variables affecting this behavior and (2) the uncertainties in the primitive variables can be described by assumed distributions (ref. 7). A multifactor interaction model (MFIM) for this relationship has been developed and incorporated in the NESSUS computer code. The MFIM is used to develop resistance probability functions as shown in figure 1, bottom right (strength).

This MFIM is applied to specific structures to probabilistically determine: (1) the resistance curve for damage (crack) initiation and (2) damage propagation and its effects on global structural response. The results for the most probable point for damage initiation, the most probable path, and the degradation in structural integrity can then be determined for specified probabilities. The important observation is that the uncertainties in damage initiation, propagation and subsequent effects on structural performance can be probabilistically simulated by the methodology described herein. It is worthy of note that this methodology, in general, is applicable to a variety of structures, and can readily be incorporated to monitor the in-service health of structures in general and aerospace structures in particular.

## RISK-COST ASSESSMENT

The methodology described previously has been extended to perform reliability and risk-cost assessments. In order to accomplish this, (1) the cost for component/structure service readiness needs to be quantified and (2) the cost as a consequence of component/structure failure must be established. Both of these have been integrated into the probabilistic structural analysis methodology (ref. 8). The results from the application of this methodology to the specific example blade are summarized in figure 3 in terms of fatigue cycles to failure.

The important observation from the aforementioned discussion is that the reliability and risk-cost of structures in general and man-rated structures in particular can be assessed using probabilistic methods of the type described herein. The implications are far-reaching because these methods are primarily computational and can be applied to existing structures to evaluate their risk for continuing service as well as those on the design board and those still in the conceptual phase.

## RELIABILITY/CERTIFICATION -- AN EMERGING APPROACH

The collective observations from the previous discussion led to an emerging approach to computationally simulate structural reliability, risk components qualification, and eventually vehicle structure certification. The general steps for this emerging approach are outlined as follows:

1. Develop a coarse structure or structural component/vehicle (global) analysis model.
2. Conduct probabilistic structural analysis (PSA) of the types described herein.
3. Identify the critical component/structure areas from the results of PSA.
4. Perform global/local PSA's to evaluate nonlinear effects and to locate probable sites of damage initiation.
5. Determine the most probable damage propagation path.
6. Evaluate probable structural degradation along this path.
7. Establish probable path extent for violation of specified structural performance criteria (for example, 10-percent increase in displacement or 5 percent reduction in the frequency of the first vibration mode).
8. Assess the corresponding reliability and risk and decide on their acceptability.
9. Schedule inspection intervals and retirement for cause criteria based on the results of items 5, 6, and 7.
10. Verify with probabilistically selected (using respective sensitivities) critical structural components and prototype structure tests.
11. Design a suitable in-service health monitoring system using the results from items 8 and 9 above in order to ascertain that the component/structure will meet the acceptable reliability and risk requirements.



## CONCLUSIONS

A methodology has been developed for the formal probabilistic quantification of uncertainties in the structural performance and subsequent reliability and risk of man-rated structures. The key elements in this methodology are: (1) probabilistic load simulation, (2) probabilistic finite element analysis, (3) probabilistic simulation of complex nonlinear material behavior, and (4) risk-cost assessment. This methodology is described in terms of fundamental aspects and application to a specific structural component which is a turbopump blade of the Space Shuttle Main Engine (SSME) and which was selected for its complexity in order to demonstrate what can be done by using this methodology. The specific example illustrates how the uncertainties in all the basic parameters (primitive variables) for loads, structure and material behavior are incorporated in order to probabilistically simulate the uncertainties in the structural response (global and local). Also, the example illustrates how the reliability and risk-cost can be assessed. Collectively, the summary of the fundamental considerations and the results from the specific example demonstrate that a formal methodology is available to evaluate the reliability and risk-cost of man-rated structures in aerospace environments as well as structures in general. In addition, an emerging approach is outlined which can be used to computationally qualify and eventually certify future structures.

## REFERENCES

1. C. C. Chamis: Probabilistic Structural Analysis Methods for Space Propulsion System Components. NASA TM 88861, 1986.
2. Anon.: AIAA/ASME/ASCE/AHS 28TH Structures, Structural Dynamics and Materials Conference, April 6-8, 1987 (papers 87-0764, -0765, -0766), pp. 252-274.
3. Anon.: Structural Integrity and Durability of Reusable Space Propulsion Systems. NASA CP 2471, 1987, pp. 117-200.
4. Anon.: AIAA/ASME/ASCE/AHS 29th Structures, Structural Dynamics and Materials Conference, April 18-20, 1988 (papers 88-2371 to 88-2376), pp. 1262-1288.
5. AIAA/ASME/ASCE/AHS/ASC 30th Structures, Structural Dynamics, and Materials Conference, April 3-5, 1989 (papers 89-1368 to 89-1372), pp. 1832-1910.
6. Anon.: AIAA/ASME/ASCE/AHS/ASC 31st Structures, Structural Dynamics and Materials Conference, April 2-4, 1990 (papers 90-1098 to 90-1102), pp. 1039-1080.
7. C. C. Chamis and D. A. Hopkins: Probabilistic Structural Analysis Methods of Hot Engine Structures. NASA TM 102091, 1989.
8. M. C. Shiao and C. C. Chamis: Probability of Failure and Risk Assessment of Propulsion Structural Components. NASA TM 102323, 1990.

TABLE 1

# HPOTP & HPFTP Parameters Phase II Engine Calculated vs Measured

Condition	HPOTP				HPFTP			
	Speed		Turbine Discharge Temp		Speed		Turbine Discharge Temp	
	Calc (rpm)	Measured (rpm)	Calc °R	Measured °R	Calc (rpm)	Measured (rpm)	Calc °R	Measured °R
Hardware — 2 $\sigma$ random	294	—	53	—	288	—	65	—
Test — 2 $\sigma$ random	210	—	157	—	398	—	20	—
Total random	360	—	165	—	554	—	70	—
Low NPSP — del	620	—	225	—	56	—	52	—
High NPSP — del	-317	—	-219	—	-94	—	-62	—
Range = random + del	1660	1500	476	400	1260	1000	114	150
Max	28090	29100	1630	1650	35742	35750	1740	1760
Nom	28100	28200	1374	1380	35130	35300	1688	1690
Min	27430	27500	1155	1250	34482	34750	1625	1610

Measured — Measured variation for phase II test set

Hardware — Variations in engine hardware

Test — Initial test conditions — Inlet temperatures & mixture ratio

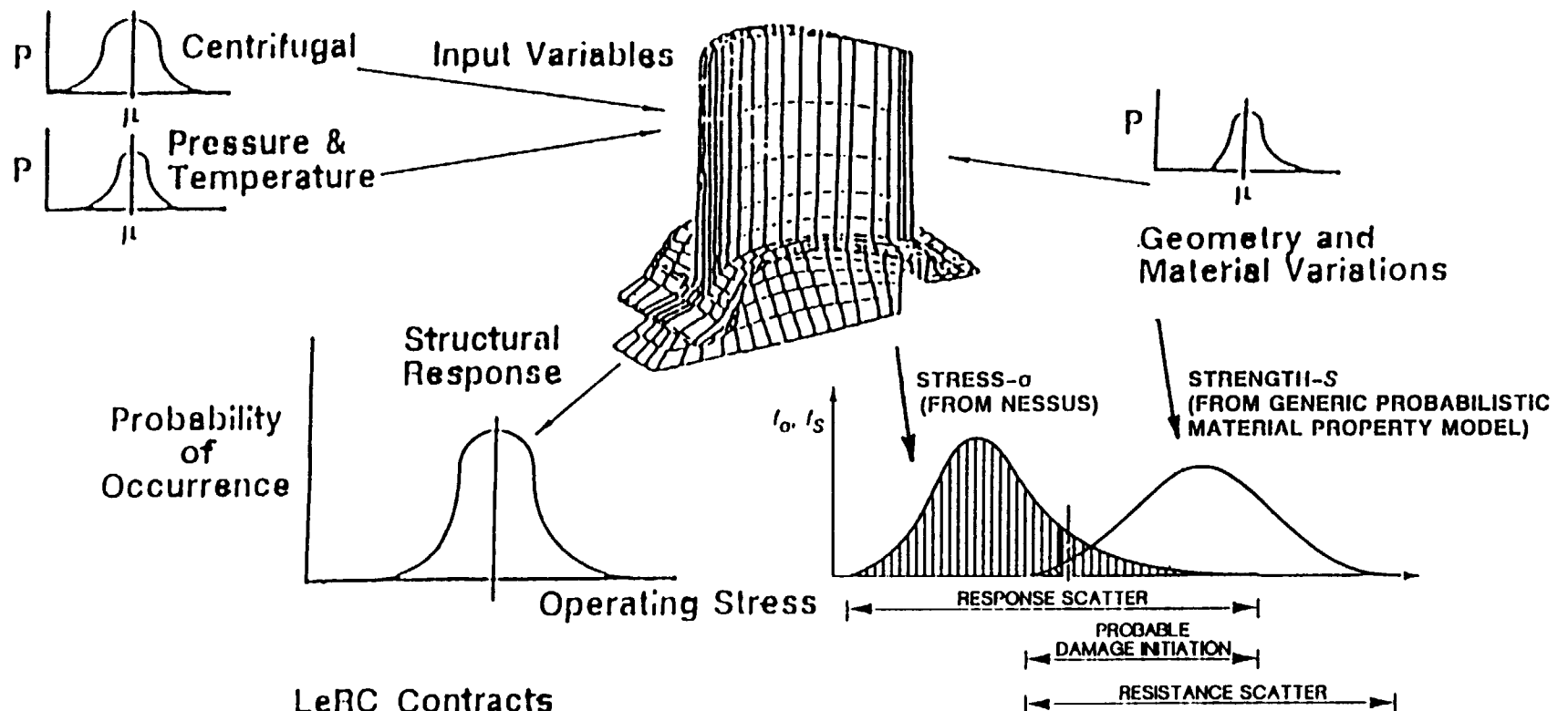
Del — Duty cycle effects of inlet pressures plus correlated 2 $\sigma$  variations of cavitation

**FIGURE 1**

# Component Response Analysis Using CLS Coupled With PSAM

Turbine Blade Loading

Nessus Turbine Blade  
Coarse Model



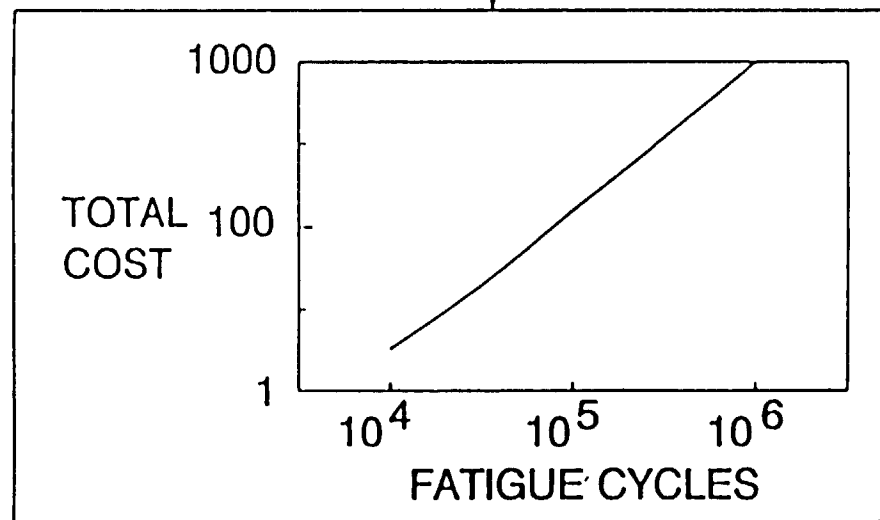
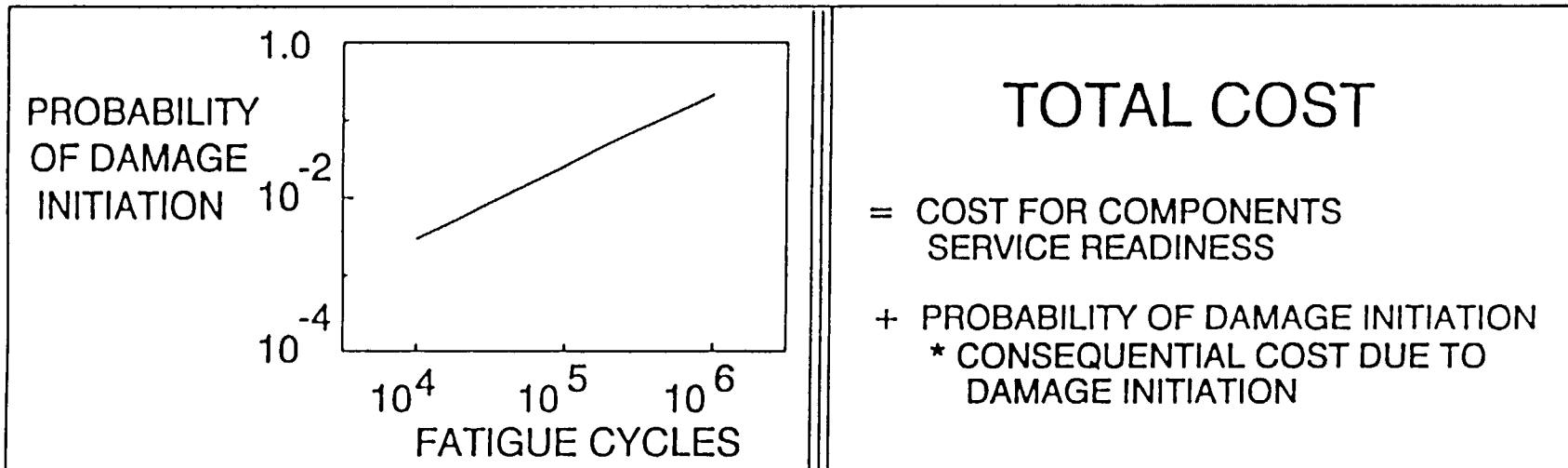
LeRC Contracts

CLS - Composite Loads Spectra

PSAM - Probabilistic Structural Analysis Methods - SWRI

FIGURE 2

# PROBABILISTIC RISK-COST ASSESSMENT



**A SEMI-AUTOMATED PROCESS FOR  
THE PRODUCTION OF CUSTOM-MADE SHOES**

**Franklin H. Farmer, Ph.D.**

**Technology Utilization and Applications Office  
NASA Langley Research Center  
Hampton, Virginia 23665-5225**

**INTRODUCTION**

A more efficient, cost-effective and timely way of designing and manufacturing custom footwear is needed. Over one-half million Americans cannot comfortably wear off-the-self, mass-produced shoes. Their requirements for custom-made shoes result primarily from congenital defects or from having worn improperly fitted footwear, but a substantial portion of these individuals have had partial amputations of the foot. However, custom-made shoes, particularly orthopedic shoes, are usually very expensive because their design and manufacture is now a highly skilled and labor intensive process; and their delivery times are often long, and getting longer, because the number of skilled shoemakers in this country is inadequate and declining rapidly.

A potential solution to this problem lies in the use of computer-aided design and manufacturing (CAD/CAM) techniques in the production of custom shoes. Although these techniques are being used in the design and manufacture of mass-produced shoes, they have only recently been applied to the production of custom shoes. In 1984, two workshops to "Establish Criteria for the Development of a Computer-Aided Design (CAD) System to be Used in the Construction and Custom Contouring of Orthopedic Footwear through Computer-Aided Manufacturing (CAM)" were sponsored by the National Institute for Handicapped Research (now National Institute for Disability and Rehabilitation Research). These workshops revealed a general consensus among the participants that the technology existed and the time was ripe for the development of a cost-effective, highly automated system which would use CAD/CAM techniques to design and manufacture custom shoes<sup>1</sup>.

In 1985, the momentum generated by those workshops resulted in the initiation of a cooperative inter-agency project to develop such a system. This project, supported equally by the Veterans Administration (VA) and NASA, has been carried out under the technical direction and management of NASA with the support of Research Triangle Institute. NASA's involvement in this project has been motivated by an opportunity to apply NASA-developed technologies, specifically the computer programs NASCAD (NASA Computer-Aided Design), RIM (Relational Information Management) and the NASA version of APT (Automatically Programmed Tools), to the solution of this problem.

Using the workshop data and recommendations as a starting point, a prototype computer-based system has been developed by a team of programmers and engineers supported by grants to North Carolina State University (NCSU) and the University of Missouri-Columbia (UMC). This system is primarily a software entity which directs and controls a 3-D scanner, a lathe or milling machine, and a pattern-cutting machine to produce the shoe last and the components to be assembled into a shoe. The steps in this process are: 1) scan the surface of the foot to obtain a 3-D image; 2) thin the foot surface data and create a tiled wire model of the foot; 3) interactively modify the wire model of the foot to produce a model of the shoe last; 4) machine the last; 5) scan the surface of the last and verify that it correctly represents the last model; 6) design cutting patterns for shoe uppers; 7) cut uppers; 8) machine an inverse mold for the shoe innersole/sole combination; 9) mold the innersole/sole; and, 10) assemble the shoe. For all its capabilities, this system still requires the direction and assistance of skilled operators, and shoemakers to assemble the shoes. Currently the system is running on a SUN3/260 workstation with TAAC application accelerator. The software elements of the system are written in either FORTRAN or C and run under a UNIX operator system.

## DETAILING THE PROCESS

Currently foot scans are being made with a commercially available 3D laser scanner, the Cyberware Model 4020. This scanner records approximately 250,000 points, which are registered in latitude, longitude and radial distance. These measurements are made with an accuracy of better than 1% and a resolution of 0.7mm. It only takes about 15 seconds per foot to acquire this data. The foot can be either in a "neutral", or partially or fully loaded position. Since the scanner can also obtain foot surface data from models made from plaster casts, this data can also be obtained on people who cannot come to the scanner location for the measurements. By January 1991, a new model scanner, designed by Cyberware specifically to acquire foot surface data, will be integrated into the system. The advantage of the new scanner is that it is more convenient to use and is also easily portable.

## CREATING A REALISTIC MODEL OF THE FOOT

### Thinning data

The high density data is needed only in those areas where the foot surface changes rapidly and thus needs to be represented in more detail, i.e. the heel and toes regions. In the other areas of the foot the changes in the surface can be effectively represented by only a few data points. A thinning routine has been developed which can relate the surface data density required to the rate of change of the surface. This routine is used to reduce the size of the data set to a minimum consistent with an accurate representation of the foot. This process minimizes the RAM storage requirements and speeds up the subsequent computations.

### Foot models

The foot surface data can be represented by any of three different models, i.e. a "point" model, a "wire" model, or a "tiled" or "shaded" model. The model selected to display the foot is a function of the process using the model. The point model is the least detailed and is thus the least realistic representation of the foot. It is used primarily for speed. When the foot is rotated from one view to another, the point model is used. The shaded model is the most realistic representation of the foot, but is the most time consuming to display. It is used when it is necessary to get a "feel" for the shape of the foot and the details of its surface. The wire model is intermediate in realism and speed.

### Rotation, color and lighting

To help provide the user with a realistic foot model, options are provided which allow shaded model rotation, coloring, and illumination. There are two rotation options: continuous and interactive rotation. The latter allows the model to be rotated among six standard views; i.e., front, back, right side, left side, top and bottom. Continuous rotation can be used to set the foot spinning on a selected axis and at a selected rotation speed. This option can be used to simulate turning a plaster foot model by hand. Color options are available for both the shaded foot model (four colors) and the background (five colors). The illumination option allows the shaded model to appear as if illuminated from any direction selected by the operator.

## INTERACTIVELY DESIGNING THE SHOE LAST MODEL

Once the foot requiring a custom shoe has been scanned, and the digitized data transferred to the computer, thinned, and presented to the operator as a foot model, the next step in the process is for the operator to interactively modify the foot model to produce a shoe last model. The shoe last modeling subsystem, called LASTMOD, provides a menu-driven user interface which allows the operator to perform the needed functional and aesthetic modifications without extensive training or familiarity with computers<sup>2</sup>. The operator can select from six different operations to achieve the modification(s) desired. These operations are "Add Volume", "Remove Volume", "Toe BuildUp", "Extend Toe", "Region BuildUp", and "Region Move". In addition to being able to view the modified model between operations, the operator can compare the current modified model with the original foot model at any point in the process. The operator can also rotate

the model and/or magnify selected areas of the model.

#### Add Volume/Remove Volume

These operations are used to add or subtract a slab of material along a previously defined cut plane, with or without a reference plane. Material below the reference plane, usually the planar surface of the foot, remains unaffected. The "Remove Volume" operation is used primarily to narrow the ankle part of the last.

#### Toe BuildUp/Extend Toes

These operations are used to either add material to the toe end of the last or to extend (stretch) the toe region of the last. This provides a toe "box" at the end of the shoe, allowing freedom of movement for the toes.

#### Region BuildUp

This operation adds material to the last in areas where the patient has an ulcer, corn, bunion, or other sensitive area that requires relief of pressure. Selectable variables include location of the entire edge of the area, the maximum thickness of the buildup, and the location of the maximum. Marks made on the foot before scanning show up on the screen in a different color and can be used as reference points.

#### Region Move

This operation moves a section of the last, usually a toe, away from the main body of the last. The space between is then automatically filled.

#### Patient Data

To aid the operator in the interactive process, a patient database feature has been included in the program. Called the "Patient Information Database", it provides a mechanism for storing and accessing data equivalent to that gathered by the practitioner (podiatrist or pedorthist) through interviews and examinations of the patient. This data is primarily actual foot measurements, related biomechanical data, and the patient's medical and orthopedic history. This database has been implemented using R:BASE for DOS (version 2.1) and RIM 5 database management systems.

### **MACHINING THE LAST**

Once the last model is complete, it is resampled with a user specified tolerance and then represented using the minimum number of Coons (parametric bicubic interpolatory) patches<sup>3</sup> to maintain accuracy requirements. The data set that represents the final model is then transferred to the Tool Path Generating and Machining Subsystem (LASTCUT) to generate cutter location data. LASTCUT generates the tool path data using one of two distinct methods: APTGESS<sup>4</sup> or SSURF<sup>5</sup>. The primary difference between these methods is that the APT-based programs generate general cutter line data which must in turn be converted by a post-processor<sup>6</sup> (UPOST) into appropriate machine-specific data, while SSURF creates tool path data directly from the shoe last model. Thus the APT-based method is slower and can have problems at interfaces between sculptured surfaces (37 patches maximum), but because the post-processor is "universal" it is applicable to any 3, 4, or 5 axis CNC milling machine tool. While SSURF is faster and can handle an almost unlimited number of Coons patches, it has no algorithms to check for tool gouging.

Before the shoe last is machined, NASCAD is used to view and, if necessary, plot the tool path. This is done to check for errors in programming or data processing. The milling of the shoe last is currently being performed on a 3-axis milling machine which has been modified by the addition of a rotary table.

## **VERIFYING THE SHOE LAST**

After a shoe last is machined, it is scanned and a computer model is generated. A verification subsystem then compares the dimensions of the computer-generated last model with those of the machined last. This subsystem checks for machining errors and decides whether the machined shoe last is within acceptable tolerances. The verification subsystem uses a similarity measure based on spine deviation, section size variation, section elongation variation, and section orientation variation. The spine is a line which passes through the centroids of a series of cross sections of the last model. The similarity measure is actually a weighted root-sum-of-squares of the differences.

## **DESIGNING AND CUTTING SHOE UPPERS**

A computer program which can be used to design and direct the cutting of shoe uppers is commercially available from Microdynamics of Dallas, TX. We are currently trying to interface the output of LASTMOD with this software. This capability will also allow the introduction of styling considerations into the final product.

## **MACHINING THE MOLD AND CASTING THE INNERSOLE/SOLE**

To produce a shoe sole with maximum fit and comfort, it was decided to cast an innersole/sole combination in a mold which reflects the plantar or bottom surface of the foot. An output file generated by LASTMOD is used to machine the mold from wood or plaster. The bottom surface of the mold is an exact reproduction of the inverted plantar surface of the foot. The orientation of that surface is variable according to the need for heel pitch (how much higher the heel will be than the toes), and the balance line (the angle the leg deviates from the vertical). The mold can also be made deeper to accommodate a shorter leg by a thicker sole. The sole mold is currently being machined on a 4-axis CNC milling machine.

There exists a wide range of materials from which the innersole/sole combination can be cast. Currently a quick-set latex material is being used for demonstration purposes. It was selected primarily on the basis of convenience, and thus other materials may be more applicable to shoe construction, wear, comfort, etc. Also a liner may be needed to separate the foot from the sole and give more comfort and better wear characteristics. The liner could easily be bonded to the inner surface of the sole during the shoe assembly operation.

## **ASSEMBLING THE SHOE**

The shoe assembly process still continues to be a manual process, but there is hope that it can be at least partially automated. This effort was outside the scope of this program, but may be addressed in any subsequent phase. At this point in the development effort conventional assembly is still being used.

## **STATUS/FUTURE PLANS**

All work towards development of a prototype system will be complete by the end of calendar year 1990. Currently a demonstration phase is underway. Two pairs of orthopedic shoes are being made using the above process. The system software and hardware will soon be turned over to the Veterans Administration for clinical testing. Once this testing is complete, a second development phase will begin which will optimize the system and make it operational. This work will probably be supported solely by the VA.

## **ACKNOWLEDGEMENTS**

This work has truly been a team effort. Each grantee has made a significant contribution to the total effort, as has each of their students. The contributions of each team member is hereby acknowledged and their areas of responsibility delineated:



- Dr. David McAllister, Professor of Computer science, Ncsu (LASTMOD, uppers design/cutting interface, project coordinator) ; J. Pietenpol, L. Harrison, S-H. Yang, D. Carver, P. Hebbar, R. Devarjan, M. Patel.
- Dr. Ren Luo, Professor of Electrical and Computer Engineering, NCSU (Last Shape Verifier, scanner interface); Y. Kim.
- Dr. Han Bao, Associate Professor of Industrial Engineering, umc (sole/innersole design and mold machining processes) ; S. Gharagozloo, J. Ainumtip, R. Aiyar, N. Lin.
- Dr. William Rasdorf, Associate Professor of Civil Engineering and Computer Science, NCSU (patient data base, NASCAD-RIM interfaces); S. Lakmazaheri, L. Boose, O. Abudayyeh.
- Dr. Ezat Sani, Assistant Professor of Industrial Engineering, NCSU (APT IV modifications and NASCAD interfaces, SSURF); H. Chou, L. Colbert, K. Srinivasan, C. Zhou.
- Mr. Robert Wallace, Industry Specialist, RTI (management support, project coordination, consultants).

Also, I would like to acknowledge the contributions of Dr. Margaret Giannini, Mr. Don Vargo, and Mr. Wijitha Ellepola, VA Central Office, Washington, DC; and Mr. Ray Gilbert, NASA Headquarters, Washington, DC.

#### REFERENCES CITED

1. Rouse, D.J.: A Proposed Design for an Orthopedic Footwear CAD/CAM Delivery System. DOE Contract #300-83-0236, 1984.
2. Pietenpol, J.L.: Computer Graphics System for the Interactive Design of Shoe Lasts. M.S. Thesis, North Carolina State University, 1989.
3. Coons, S.A.: Surfaces for Computer Aided Design of Space Forms. MAC-TR-41 Project MAC, M.I.T., 1967.
4. Srinivasan, H.: APT Program Generator for the Computer Aided Manufacture of Custom Orthopedic Footwear. M.S. Thesis, North Carolina State University, 1989.
5. Colbert, L.: SSURF - A Tool Path Planning System for Machining Three-Dimensional Sculptured Surfaces. M.S. Thesis, North Carolina State University, 1989.
6. Chou, H.L.: APT Universal Postprocessor for Multi-Axis CNC Milling Machine Tools. M.S. Thesis, North Carolina State University, 1989.

## LIGHTWEIGHT, FIRE-RETARDANT, CRASHWORTHY AIRCRAFT SEAT CUSHIONING

Leonard A. Haslim  
Paul T. McDonough

Civil Technology Office  
NASA Ames Research Center

The primary goal of NASA Ames Research Center's Civil Technology Office (CTO) is to provide American industry with hi-technology products. The research and development done at Ames is transferred smoothly and expediently so as to keep America on the cutting edge in a competitive global marketplace. The majority of the ideas and products developed by the CTO are aerospace-related, but often spin-offs from the technology reached non-aerospace markets. One such technology is the light-weight, fire retardant, crashworthy aircraft seat cushioning which promises to revolutionize the safety and comfort of both aerospace and non-aerospace seating.

Presently, most aircraft, automobile, train, bus, hotel and hospital seating is composed of polyurethane foam. Although it is inexpensive and comfortable, the foam is lethal when ignited. It burns quickly, producing a dense white smoke and emits highly lethal hydrogen cyanide gas, if not immediately ventilated, will cause death. In 1977 in the Canary Islands, two Boeing 747's collided on the ground. Most of the 583 passengers died from smoke inhalation and burns. In many crash investigations, it was found that the smoke, not the fire, was the lethal factor. In the hopes of avoiding such catastrophe, the Civil Technology Office began experimenting with new materials and seat designs which would alleviate the use of the urethane foam.

The Safety Seat Cushioning (SSC) was designed for both safety and comfort. Composed of advanced fabric reinforced composites, it is lightweight, fire-retardant and crashworthy. The seat design consists of central elliptical tubular spring supports made of fire-resistant and fatigue-durable composites surrounded by a fire-blocking sheath. The cushioning is made crashworthy by incorporating energy-absorbing, visco-elastic layers between the nested, elliptical-hoop springs; a highly desirable feature for helicopters. The design is intended to provide comfortable seating that meets aircraft-loading requirements without using the conventional polyurethane materials.

Aside from its fire-resistant and energy-absorbing characteristics, the SSC is lightweight, economical, simple to fabricate, structurally strong and easily maintained. Although its initial intent was for aircraft and helicopter seating, the SSC has also become popular with automotive and furniture manufacturers.

Several key features of the SSC have attracted the automotive industry. For the same safety concerns as with aircraft seats, automobile accidents would be much less hazardous without the combustible and toxic urethane foam seating. Also, the durability and maintainability of the hoops will provide long-lasting comfort to the occupants. Another interesting feature is that heating ducts can be inserted into the SSC tube assembly. Because of its energy-absorbing characteristic, a replacement for a five mile an hour bumper has also been proposed.

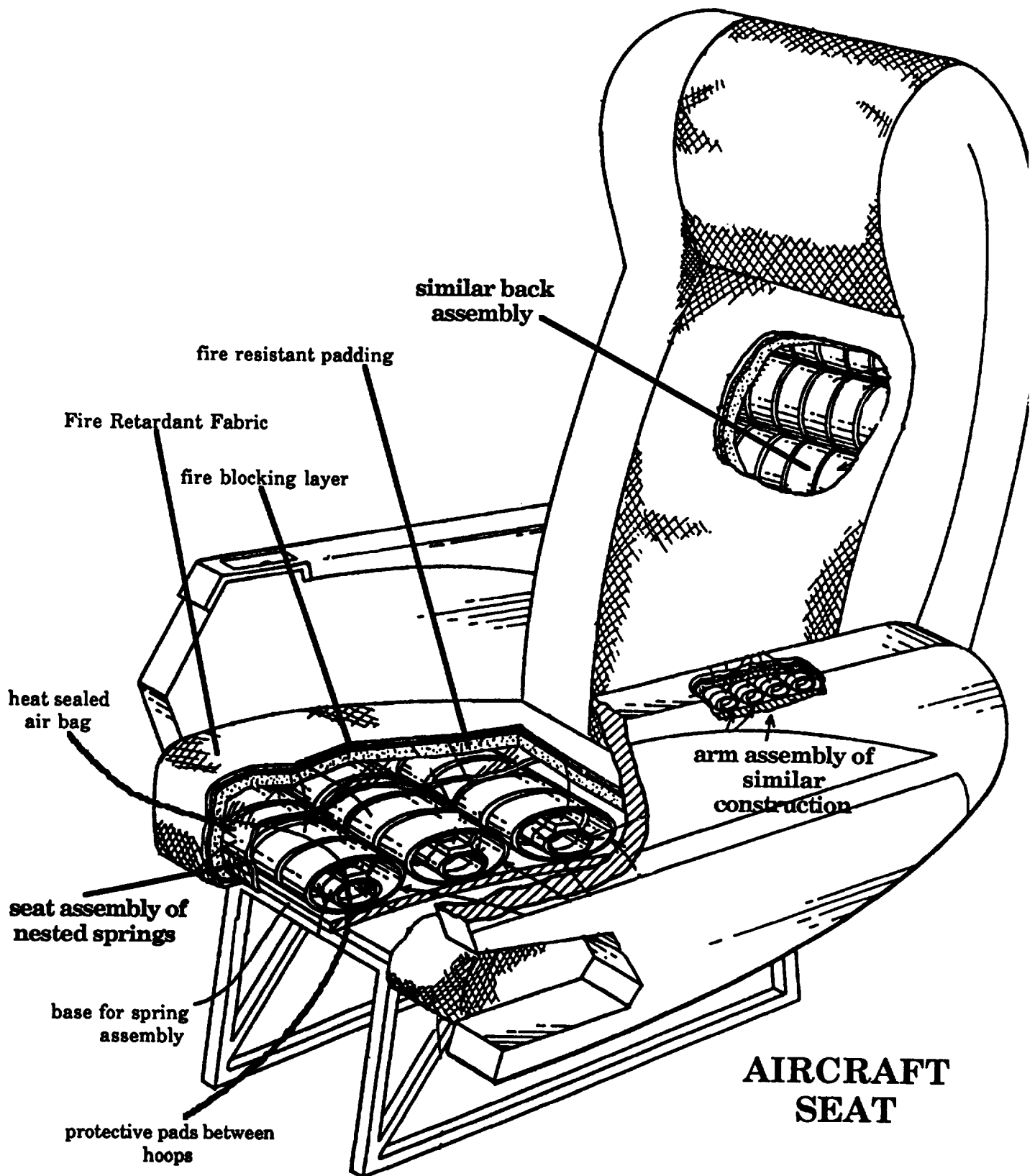
The SSC can be applied to most mass transit systems, particularly busses and trains. Train and bus passengers will benefit from the SSC in both safety and comfort. In the 1989 Kentucky school bus tragedy, three adults and 24 children were killed. All of the victims died of smoke inhalation, not from injuries sustained in the crash. Once again, polyurethane foam's toxic and flammable properties inhibited escape and rescue.

The Civil Technology Office has also been in contact with navy ship manufacturers who want to incorporate the SSC into ship design. Not only for its fire-retardant properties, but its energy-absorbing

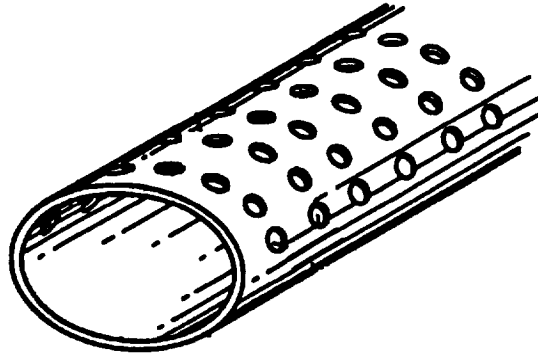
characteristic will help reduce personal injury and damage to ship computer hardware caused by explosive mines and near misses. In such a scenario, the ship is rocked with a violent jolt which can cause spinal injuries and destroy sensitive electronic systems.

Other non-aerospace applications include home, hotel and hospital furniture and bedding. Furniture manufacturers are also interested in the SSC's durability and comfort aspects, as well as its fire-resistance. The incidence of injury in high rise hotel fires would be greatly reduced if the polyurethane foam bedding and furniture were eliminated. Escape, rescue and eventual extinguishing would all be enhanced. Hospital bedding could be made more comfortable for patients with the use of heating ducts applied in a therapeutic manner.

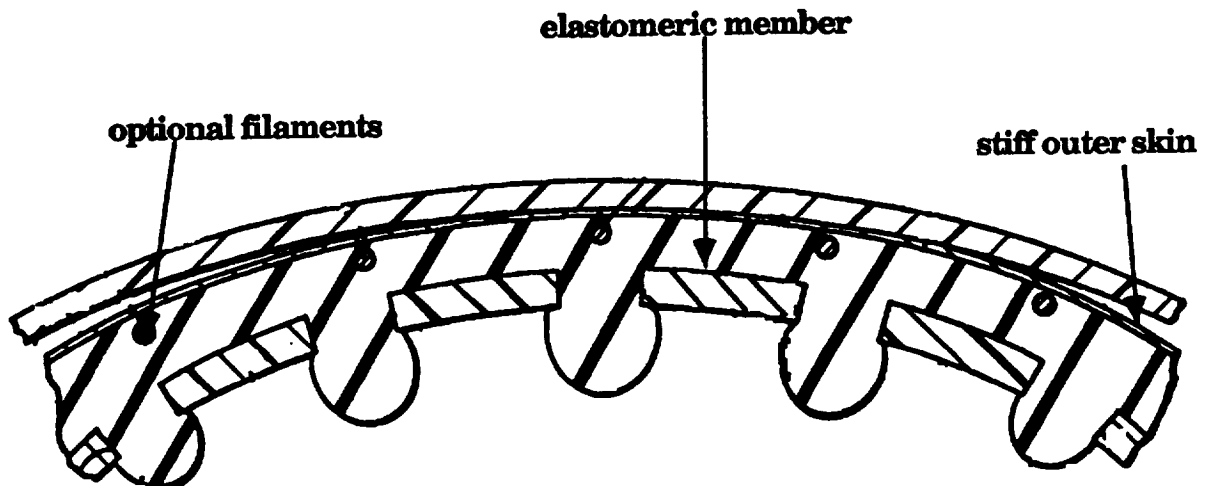
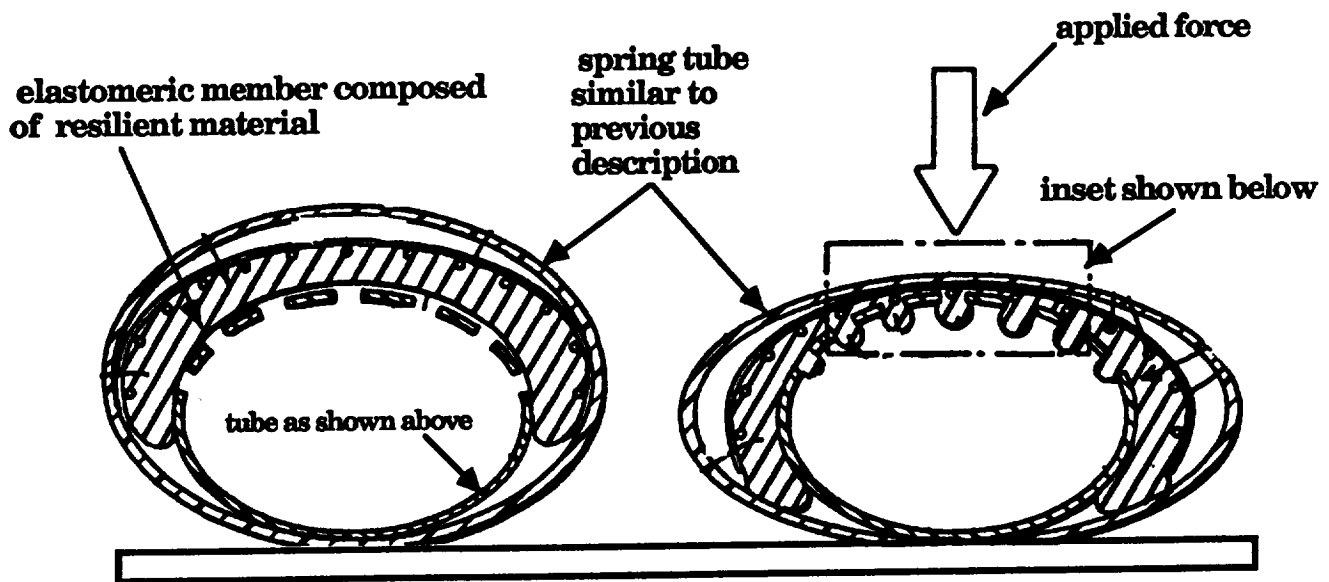
The applications of the SSC are varied and numerous. The manufacture and cost of the seat spring assembly is both simple and economical. Patented by NASA and easily accessible to industry through the Technology Utilization Office, the Civil Technology Office's SSC is a fine example of NASA technology developed with safety, comfort and well-being in mind.



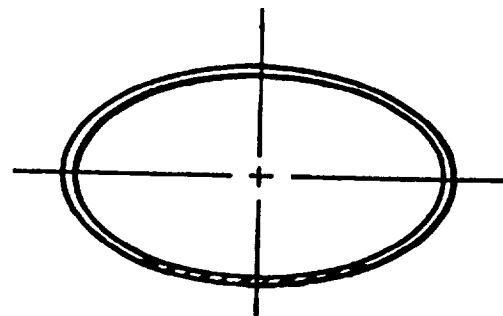
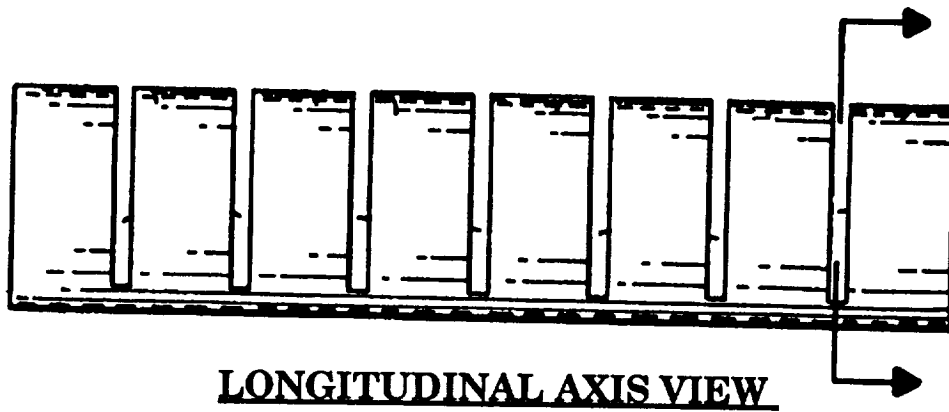
## **NESTED SPRINGS USED WITH INNER ELASTOMERIC LINING**



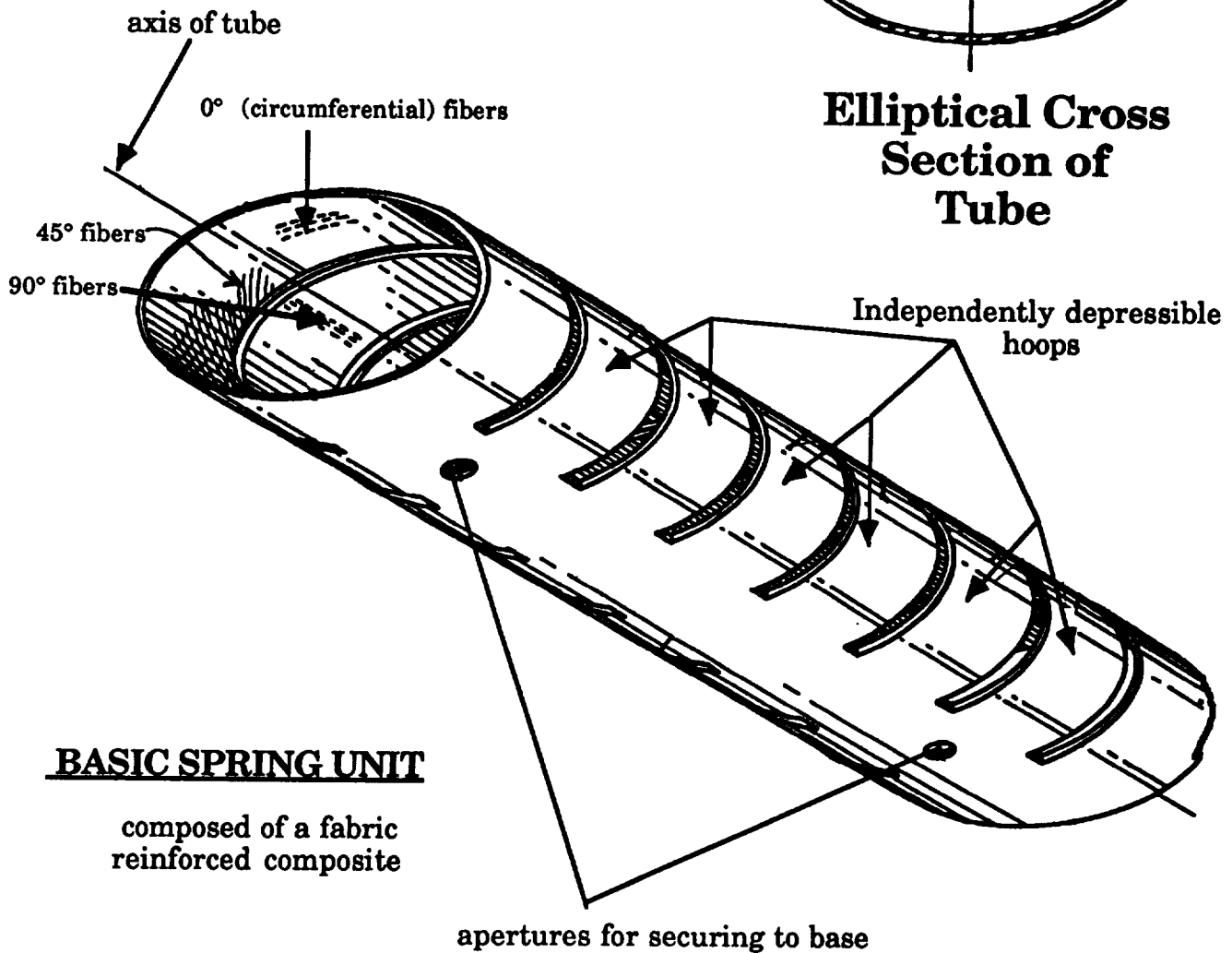
### **CONSTRUCTION OF INNER NESTED TUBE**

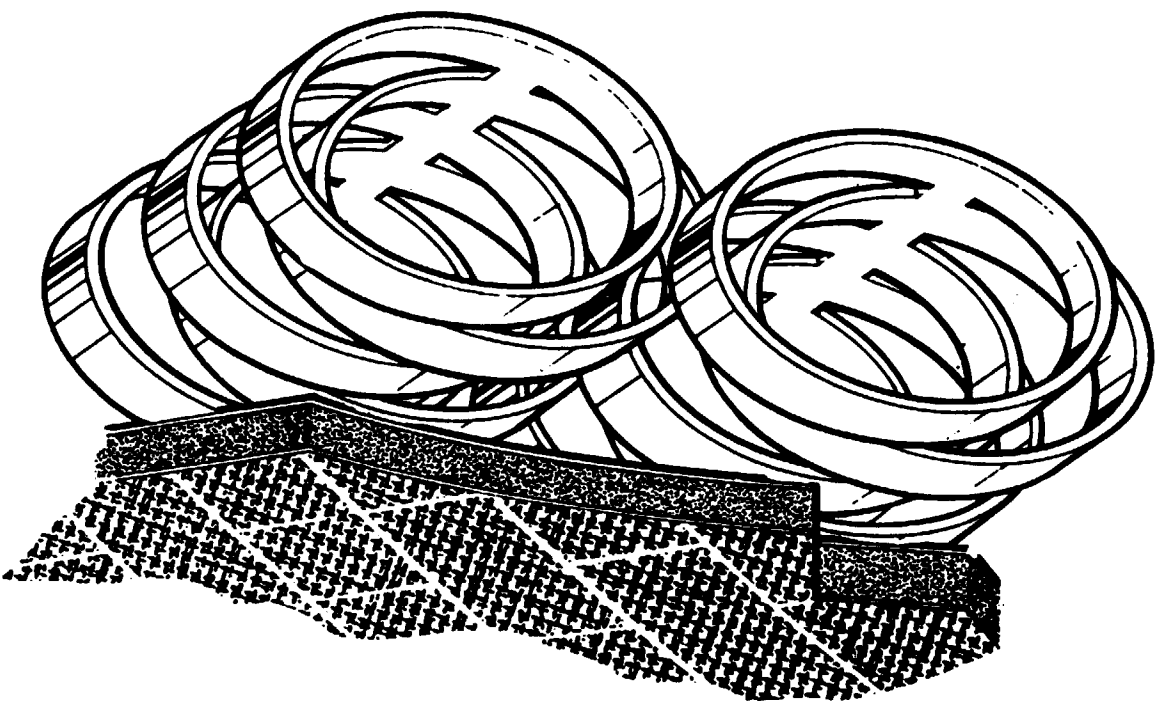


**Applied force is damped as elastomer flows through inner nested tube**



**Elliptical Cross  
Section of  
Tube**









## **SESSION F - POWER, ENERGY AND CONTROL SYSTEMS**

**Tuesday November 27, 1990**

- **Civil Air Transport: A Fresh Look At Power-By-Wire And Fly-By-Light**
- **The Free-Piston Stirling Engine - From Space Technology To Terrestrial Applications**
- **Solar-Powered Stirling Cycle Electricity Generator**
- **Four Quadrant Control Of Induction Motors**
- **High Power Bi-Polar Lead-Acid Batteries**
- **Scientific CCD Technology At JPL**
- **Advanced Thermal Technology For Commercial Applications**

**PRECEDING PAGE BLANK NOT FILMED**



## CIVIL AIR TRANSPORT: A FRESH LOOK AT POWER-BY-WIRE AND FLY-BY-LIGHT

Gale R. Sundberg  
National Aeronautics and Space Administration  
Lewis Research Center  
Cleveland, Ohio 44135

Abstract

Power-by-wire (PBW) is a key element under subsonic transport flight systems technology with potential savings of over 10 percent in operating empty weight and in fuel consumption compared to today's transport aircraft. The PBW technology substitutes electrical actuation in place of centralized hydraulics, uses internal starter-motor/generators and eliminates the need for variable engine bleed air to supply cabin comfort.

The application of advanced fiber optics to the electrical power system controls, to built-in-test (BIT) equipment, and to fly-by-light (FBL) flight controls provides additional benefits in lightning and high energy radio frequency (HERF) immunity over existing mechanical or even fly-by-wire controls. This paper will review the program plan and give a snapshot of the key technologies and their benefits to all future aircraft - civil and military.

Introduction

NASA in response to a directive from the U.S. Senate has developed a multiyear technology development and validation plan that will help the United States retain its leadership in aeronautics research and technology and compete in the International marketplace for future civil aircraft. One of the key elements under subsonic transport flight systems technology is power-bywire (PBW) with potential savings of over 10 percent in operating empty weight and fuel consumption compared to today's aircraft (Ref. 1). Fly-by-light (FBL), which is the replacement of electronic data transmission, mechanical control linkages, and electronic sensors with optical components and subsystems, is another key element.

In the NASA Lewis Research Center study reported in Ref. 1 the stated benefits are shown to be possible through the integration of an advanced secondary electrical power system into a civil transport aircraft using a Boeing 767 as a baseline. While the particular benefits may depend on aircraft size and type, engines, and specific electrical and flight control systems, the range of improvements is consistent with available advanced technologies.

The primary weight reduction occurred in the secondary power system when the baseline hydraulic, pneumatic and electrical subsystems were replaced with a single, advanced high frequency, sinusoidal power management and distribution (PMAD) system with controlled energy flow and load management. An advanced fly-by-wire flight control system using electrical actuators and advanced low fixed bleed high bypass ratio engines contribute the remaining major advantages according to the NASA Lewis study. The single electrical power system provides higher component utilization with significant reductions in parts count, weight, failure modes, and cost of ownership.

The PBW technology eliminates the need for hydraulic actuation and for engine bleed air to supply cabin comfort and anti-icing. It enables integral starter/generators for engine starting and power generation to be used with advanced energy-efficient engines without gearboxes. These changes significantly improve fuel efficiency and reduce aircraft weight. Improved safety and dispatch reliability combined with lower maintenance and direct operating costs are additional tangible benefits.

According to the study, eliminating the engine bleed for powering the environmental control system (ECS) and for anti-icing provided the largest single fuel savings of any proposed change. An electric motor driven ECS with electric impulse driven de-icers are the proposed replacements.

Subsequent to the NASA Lewis study, additional technologies have emerged. The application of advanced fiber optics to the electrical power system controls, to built-in-test (BIT) equipment, and to fly-by-light (FBL) flight controls provides additional benefits in lightning and high energy radio frequency (HERF) immunity over existing mechanical or even fly-by-wire controls.

This paper will review the program plan and take a fresh look at some of the key technologies and their benefits to all future aircraft - civilian and military. Figure 1 shows the payoffs and the major tasks planned for the PBW element. Figure 2 summarizes the primary elements of the FBL/PBW 5-yr program plan under the NASA Subsonic Transport Initiative.

### The Electrical Powerplant

The heart of the electrical power system is a multiredundant, fault tolerant, microprocessor controlled, power management and distribution (PMAD) system. It incorporates bidirectional inverters driven through a high frequency, resonant utility bus connected to internal starter-motor/generators, thereby eliminating gearboxes and the need for separate APU's and engine starters. The high frequency utility bus permits all the advantages of ac for stability, fault clearing, differential monitoring and control, as well as significant crew/technician safety because of its low energy per pulse nature and ease of ground fault interruption. Figure 3 shows a typical utility bus architecture for an ac PMAD system using advanced electronic switching for power conditioning, control, distribution, protection and a fault tolerant architecture (Ref. 1).

The characteristics of the advanced PMAD system provide the conduit to the benefits for PBW in civil transport and perhaps certain military aircraft. The resonant, high frequency (>10 kHz) link drive enables either multiphase low frequency ac sources or dc sources to operate at their optimum voltage and frequency at the input. Since the main inverters switch at the zero crossing of either current or voltage, they minimize power losses, component stress, EMI/EMC and the need for heavy, bulky filters. The high frequency bidirectional conversion and synthesis significantly reduces the size and mass of the electronics components, controls and systematics.

The key benefit, however, comes in the ability to drive all kinds of motors (including rugged induction motors with high temperature capability) in either direction with independent control of torque, speed and maximized, efficient operation over the entire speed range. In effect, all load control including variable speed motor drive is accomplished by sorting and steering the high frequency sinusoidal pulses to the appropriate power switch, motor winding or energy storage element. Voltage regulation, power quality and energy flow are determined and managed at all times and within specified limits. Multiple levels of redundancy are easily accommodated in the system, providing fault containment, fault recovery, and maximized end-to-end efficiency.

### Electrical Actuation

Existing electrical actuation technology in the 5- to 50-hp range will be adapted to the flight control and other actuation requirements on an aircraft. Prototype electrical actuators operating from the distributed power bus will be built and demonstrated in a full avionics control environment.

The electric actuators would replace hydraulic actuators, servovalves, and mechanical control linkages. An advanced electrical actuation system includes electromechanical and/or electrohydraulic actuators, load receivers, redundant digital data buses, and remote terminals. Electrical actuators perform the same functions as hydraulic actuators with lower weight, higher efficiency and without sizing restrictions. (Ref. 2)

The digital data bus and electro-optical sensors and controls promise to be a much lighter weight, more reliable and EMI immune approach than fly-by-wire or mechanical controls. Also, many functions

previously done with hardware can now be done with software. This may include such functions as control surface damping and trim.

### Status and Health

Microchip level BIT will be built into the hardware to provide a "fingerprint", which may include component characteristics, test information and validation parameters. Such smart BIT chips could provide health self-testing for pre-flight checkout, for in-flight status and for maintenance assistance and records.

The keys to autonomous, growable power and control systems are simple, smart, replicative logic structures. Pushing the intelligence down to the power switch and circuit level enables easy verification, validation, status, and maintainability. It provides step-by-step transitions from manual to autonomous controls. Integrated health monitoring, incipient fault prediction and a controlled evolution of power and avionics systems are readily accommodated with each node communicating with other nodes via simple, common words. This enables distributed intelligence for fault containment, fault tolerance, and autonomous control without massive software investments.

### Electro-optical Controls

The fly-by-light (FBL) controls are proposed as a replacement of electrical data transmission, mechanical control linkages and electronic sensors with optical components and subsystems. They circumvent electromagnetic interference (EMI) concerns in applying digital controls by providing lifetime immunity to signal EMI without need for shielding. The FBL technology will demonstrate optical sensors and interfaces with improved lightning and HERF immunity.

The FBL program will identify, develop, and evaluate an optical sensor suite. The sensors will be integrated into innovative electro-optical based fault-tolerant architectures using optical networks and multiplexer/demultiplexer techniques. Performance and reliability assessments of the fault tolerant processor and architectures will provide a basis for developing hardware and software for flight test and in-service evaluation.

Several ongoing activities will feed technology into the FBL program: the fiber optical control system integration (FOCSI) program, the optical propulsion management interface system being designed into the advanced transport operating system aircraft, and a fiber optical transmitter/receiver with a dc 4-GHz bandwidth. An extensive data base and experimental investigation of lightning effects on digital electronics will serve as a baseline for assessing FBL enhancements to the flight control system.

### Summary

The goal of the FBL/PBW program is to accomplish credible flight tests and demonstration of full authority, all digital FBL/PBW transport aircraft systems. Performance will be evaluated in the stress of flight environments. The program will use the NASA advanced transport operating system aircraft. Using parallel operation of experimental equipment on the basic aircraft will maximize flight test safety.

The flight tests will be designed to verify and evaluate the integrated system. FAA participation and coordination will be integral to developing a prototype certification model.

### Reference

1. Hoffman, A.C. et al., Advanced Secondary Power System for Transport Aircraft, NASA TP-2463; May, 1985.
2. Sundberg, G. R., Advanced Launch System (ALS): Electrical Actuation and Power Systems Improve Operability and Cost Picture, NASA TM-102547; May, 1990.

# POWER-BY-WIRE TECHNOLOGY

HERE IS WHAT WE PLAN TO DO

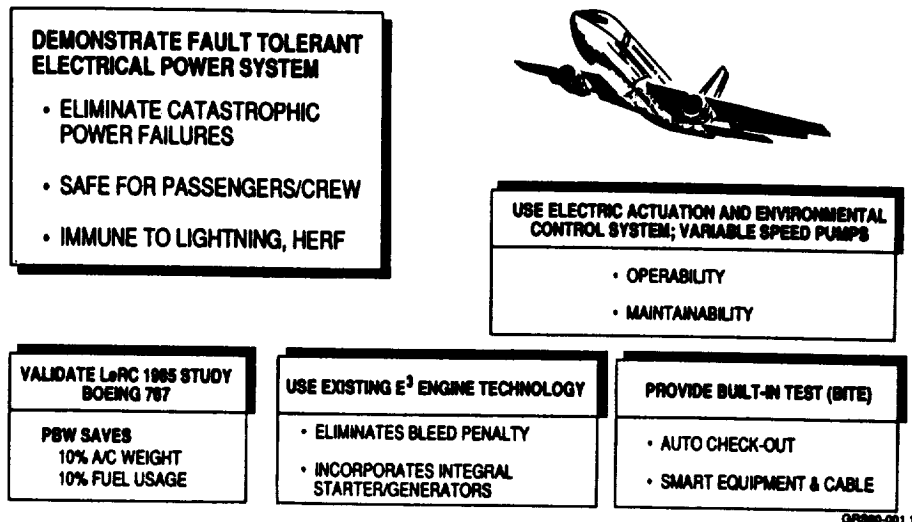


FIGURE 1. - MAJOR TASKS AND BENEFITS OF POWER-BY-WIRE TECHNOLOGY.

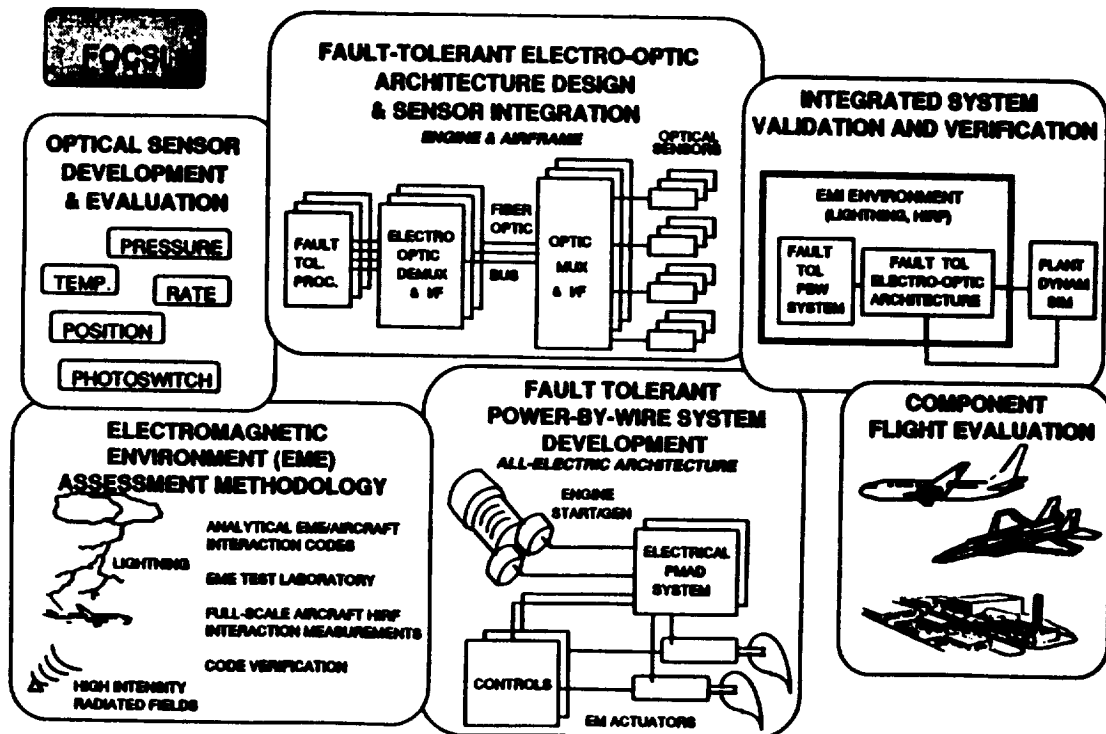


FIGURE 2. - ELEMENTS OF FLY-BY-LIGHT/POWER-BY-WIRE PROGRAM.

ORIGINAL PAGE IS  
OF POOR QUALITY

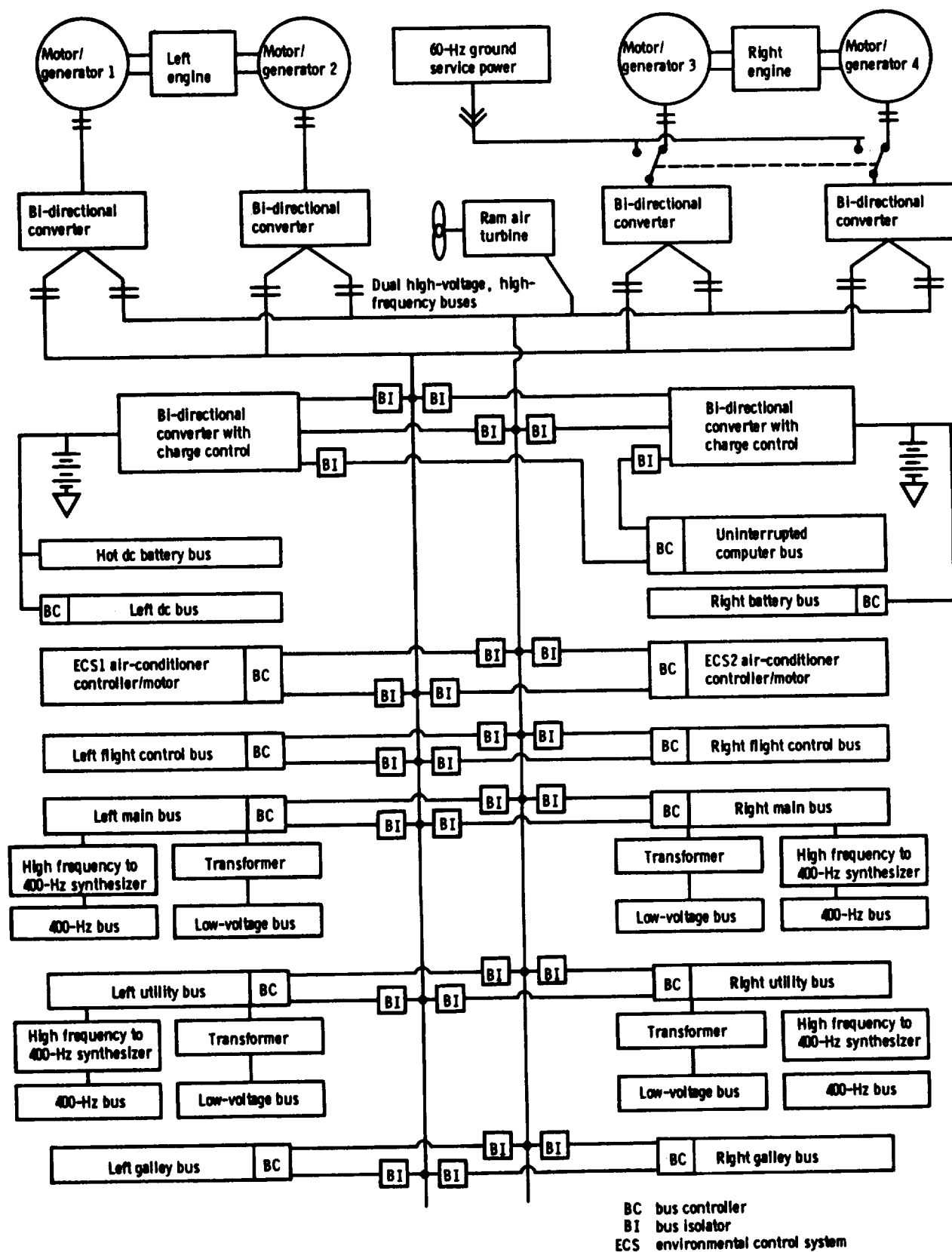


FIGURE 3. - ADVANCED POWER DISTRIBUTION SYSTEM.

ORIGINAL PAGE IS  
OF POOR QUALITY

**THE FREE-PISTON STIRLING ENGINE - FROM SPACE TECHNOLOGY  
TO TERRESTRIAL APPLICATIONS**

**James Dudenhofer, Chief  
Stirling Technology Branch  
NASA Lewis Research Center  
Mail Stop 301-2  
Cleveland, OH 44135**

**DOCUMENTATION OF THIS PAPER WAS NOT PROVIDED FOR INCLUSION IN THESE  
PROCEEDINGS. FOR FURTHER INFORMATION, PLEASE DIRECT ALL INQUIRIES TO THE  
NAME AND ADDRESS LISTED ABOVE.**



## SOLAR POWERED STIRLING CYCLE ELECTRICAL GENERATOR

Richard K. Shaltens

NASA Lewis Research Center  
Cleveland, Ohio

## ABSTRACT

Under NASA's Civil Space Technology Initiative (CSTI) the NASA Lewis Research Center is developing the technology needed for free-piston Stirling engines as a candidate power source for space systems in the late 1990's and into the next century. Space power requirements include high efficiency, very long life, high reliability, and low vibration. Further, system weight and operating temperature are important. The free-piston Stirling engine has the potential for a highly reliable engine with long life because it has only a few moving parts, non-contacting gas bearings, and can be hermetically sealed.

These attributes of the free-piston Stirling engine also make it a viable candidate for terrestrial applications. In cooperation with the Department of Energy, system designs are currently being completed that feature the free-piston Stirling engine for terrestrial applications. Industry teams have been assembled and are currently completing designs for two Advanced Stirling Conversion Systems utilizing technology being developed under the NASA CSTI Program. These systems, when coupled with a parabolic mirror to collect the solar energy, are capable of producing about 25 kW of electricity to a utility grid. Industry has identified a niche market for dish Stirling systems for worldwide remote power application. They believe that these niche markets may play a major role in the introduction of Stirling products into the commercial market.

## INTRODUCTION

The NASA Lewis Research Center, in cooperation with the Department of Energy (DOE), Solar Thermal Technology Program, is currently working with contractor teams designing solar Stirling systems which feature the free-piston Stirling engine for terrestrial applications. Recent studies have concluded that solar dish electric systems utilizing reflux receivers integrated with Stirling engines are a promising candidate as future power sources for remote power applications and ultimately for the utility industry. The free-piston Stirling technology currently being developed by NASA for space applications has the potential to meet DOE's long term goals for performance and cost. (Ref 1)

The Stirling Technology Branch at the NASA Lewis Research Center is responsible for a variety of projects which feature the free-piston Stirling engine. Work is being performed to develop the necessary technologies for free-piston Stirling conversion systems under NASA's Civil Space Technology Initiative (CSTI). Further discussion and description of the NASA CSTI program can be found in reference 2. This paper will overview the NASA managed Advanced Stirling Conversion System (ASCS) project which has the potential to meet DOE's performance and long term cost goals.

## FREE-PISTON STIRLING ENGINE

The free-piston Stirling engine was invented in 1962 by William Beale at Ohio University. Beale's early work resulted in small-scale fractional-horsepower free-piston Stirling engines which demonstrated the basic operating principles. The major advantage of the free-piston Stirling engine over the kinematic Stirling engine is that it has only a few moving parts (a displacer piston and power piston),

can use noncontacting gas bearings and can be hermetically sealed, thereby increasing the potential for high reliability and very long life. Free-piston engines have no mechanical mechanism coupling the reciprocating elements to each other, rather they are coupled through the forces exerted by the working fluid. The engine will resonate at a frequency determined by the combined dynamics of the piston and displacer. A simplified drawing of the free-piston Stirling engine integrated with a linear alternator is shown in figure 1.

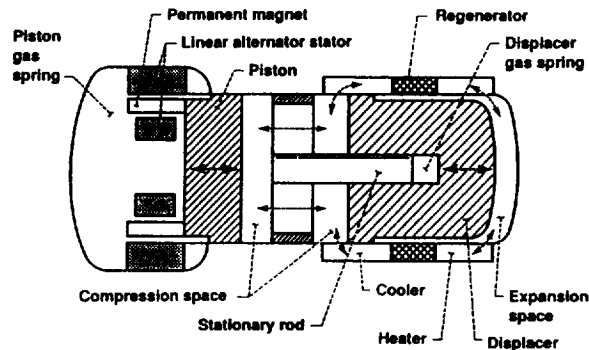


FIGURE 1. - FREE-PISTON STIRLING ENGINE WITH LINEAR ALTERNATOR.

CD-95-40458

A detailed discussion of the free-piston Stirling engine is contained in reference 3. Only a few organizations in the United States are currently developing free-piston Stirling technology. These include Sunpower Inc. of Athens, Ohio, Mechanical Technology, Inc. (MTI) of Latham, New York, Stirling Technology Company (STC) of Richland, Washington, DOE Oak Ridge National Laboratory and NASA Lewis. Free-piston Stirling engines have been designed and built in power levels ranging from fractional kW up to 25 kW (33 hp). By 1982 a few free-piston engines had been built in the 3- to 4-kW size range. The NASA space power demonstrator engine was designed and built in less than 18 months by MTI, and had delivered more than 22 kW of output power by 1986. Currently, power output of the MTI design is in good agreement with the original predictions at a temperature ratio of 2. Applications of free-piston Stirling systems range from a small implantable heart pumps, to generator sets, heat pumps and cryocoolers.

The major differences between the kinematic and free-piston Stirling engines for the solar application are in the drive and gas systems. The mechanical drive system components (drive shaft, bearings, oil pump, piston rings, crossheads and shaft seals, etc.) and the gas system (the power control system) components (compressor, check valves, storage tank and power control valve, etc.) have wear potential which would reduce engine life in the kinematic configuration. Design life of up to 20 000 hr may be obtained for reciprocating engines such as the kinematic Stirling when modified for a constant speed and load. The free-piston Stirling engine, however has no significant side loads, which minimizes wear mechanisms and allows for long life. The use of noncontacting gas bearings during operation should permit the free-piston conversion system to exceed the 60 000 hr life requirement for the solar application without difficulty. In addition, the free-piston conversion system can be hermetically sealed therefore eliminating the need for a gas makeup system. Further, the free-piston Stirling is oil free, eliminating the problem of the dynamic seals life and concern of regenerator contamination. Analysis has shown the same high brake efficiency as a percentage of Carnot efficiency is expected from either the free-piston or kinematic Stirling configurations. The expected high efficiency, along with the inherently simpler design, make the free-piston Stirling the engine of choice for the long term solar application. These attributes of the free-piston Stirling engine also make it a viable candidate for several terrestrial applications.

Recently Cummins Power Generation (a subsidiary of Cummins Engine Company), of Columbus,

Indiana has teamed with Sunpower and Westinghouse Electric Corporation, Pittsburgh, Pennsylvania, has teamed with STC in efforts to commercialize the free-piston Stirling conversion systems. Free-piston Stirling engines are currently under development for the solar terrestrial power application. Both Cummins and Westinghouse have identified a niche market for dish Stirling systems for the worldwide remote electric power applications (Refs. 4 and 5). Power levels required range from 5- to 25- kW. Both organizations believe that successful operation in these niche markets may play a major role in the introduction of Stirling products into the commercial marketplace.

## **ADVANCED STIRLING CONVERSION SYSTEMS**

NASA Lewis is providing management of the ASCS project through a cooperative interagency agreement with the Department of Energy. Cost shared contracts are in place with contractor teams which include manufactures to enhance the free-piston Stirling technology and subsequent commercialization of the ASCS. Contractor teams headed by the Cummins Engine Company and the Stirling Technology Company/Westinghouse completed the preliminary design of each ASCS in late 1989. (Ref. 6)

Each "system" consists of a solar energy receiver, a liquid metal heat transport system, a free-piston Stirling engine, the engine heat rejection system, an alternator or generator either directly or indirectly coupled to the utility grid, and the appropriate controls and power conditioning. The preliminary design completed by the Cummins Team features a heat pipe receiver integrated with a free-piston Stirling engine/linear alternator conversion system (Ref. 7). The STC preliminary design features a reflux boiler receiver integrated with a Stirling engine/hydraulic conversion system (Ref. 8). Each ASCS is designed to mount on, and to receive concentrated solar energy from an 11 m test bed concentrator (TBC) located at the Sandia Test Facility in Albuquerque, New Mexico. The ASCS Project design requirements and the Sandia TBC characteristics are given in reference 9.

DOE's requirements for high efficiency along with long life and high reliability makes the free-piston Stirling engine an ideal candidate for the terrestrial application (Ref. 1). Although the duty cycles for the space and terrestrial applications are quite different, the key technologies are similar. Work is ongoing for the demanding materials requirements which include materials characterization along with and life and reliability predictions for the liquid metal heat transport system and Stirling heater head. Use of noncontacting gas bearings during operation should permit the free-piston Stirling engine to meet or exceed the 60 000 hr life requirement for the solar application. In addition, the conversion system can be hermetically sealed therefore eliminating the need for working gas makeup system typical of kinematic engine systems.

The DOE cost goals are shown in Table I. The receiver and conversion system costs have been combined to provide a total system cost for the ASCS, excluding the concentrator. The ASCS total cost is based on collecting concentrated solar energy into a receiver from an 11 m parabolic dish while providing 25 kW or more electrical power to a utility grid. Based on an independent assessment, both ASCS concepts, have the potential to meet the DOE long term cost goal of \$452/kW<sub>e</sub> (Ref. 10). A comparison of the Cummins and STC preliminary design is provided in Table II. The expected high efficiency, along with inherently simple design, and potential for lower cost, make the free-piston Stirling the engine of choice for the long term solar application.

## **CUMMINS ASCS PRELIMINARY DESIGN**

### Cummins Preliminary Design Team

Cummins Power Generation (CPG), a subsidiary of Cummins Engine Company (CEC), Columbus, IN

was responsible for project management and system integration of the free-piston linear alternator concept into the ASCS preliminary design. The Cummins ASCS is shown in Fig. 2. CPG teamed with Thermacore Inc., Lancaster, PA for the heat transport system, and Sanders Associates, Nashua, NH, for the solar receiver. Further, Sunpower Inc., Athens, OH provided the analytical and design experience for the free-piston Stirling conversion system. Cummins Electronic Company, Columbus, IN was responsible for the system controls and power conditioning, while McCord Heat Transfer Corp., Wall Lake, MI, for the external cooling system. Cummins R&D Engineering, Columbus, IN provided expertise for the FMEA, materials and manufacturing technology. Consultants used during the preliminary design process included: Onan Corporation, Minneapolis, MN, (a subsidiary of Cummins) for alternator manufacturing issues and Dr. F. Demofte (University of New Mexico) for gas bearing analysis.

#### Receiver/Heat Transport System (HTS)

During the conceptual design phase of the ASCS Project, a heat pipe concept was selected to be integrated with the Stirling conversion system. Advantages cited for use of the heat pipe included a small liquid metal inventory along with demonstrated design theory. Concerns identified for the heat pipe technology are high construction cost and the vulnerability of the wicking system to dryout (a single point failure) and subsequent burnout of the receiver.

The receiver/HTS is a single heat pipe on a hemispherical shell as shown in Fig. 3. The heat pipe evaporator is designed to operate continuously, providing heat to the Stirling engine at 700 °C. A nominal insolation of 950 W/m<sup>2</sup> provides about 75 kW of thermal energy while a maximum peak insolation of 1100 W/m<sup>2</sup> provides about 87 kW of thermal energy to the absorber surface. The absorber is a full 180 degree hemisphere with a sintered powder metal wick on the evaporator surface. Inconel 625 was selected as the base material for the absorber and the body of the HTS. The working fluid is high purity sodium. The surface of the evaporator is covered with both a circumferential and a radial artery system. The manufacturing cost at a rate of 10 000 units per year is estimated at \$844 in 1984 dollars.

#### Engine/Linear Alternator System

The Stirling conversion system is a single cylinder free-piston Stirling engine integrated with a linear alternator to directly convert the power generated to electrical energy for the grid. Figure 4 shows a cross section of the engine/linear alternator conversion system. The heater head is designed at operation at 700 °C with an engine operating frequency of 60 Hz. The working fluid is Helium at a mean pressure of 10.5 MPa and the regenerator and cooler have an annular configuration. A single magnet configuration is used for the linear alternator along with internal magnetic springs for centering of the piston. Hydrodynamic gas bearings were designed to allow the use of non-contacting seals which eliminate wear mechanisms during operation. Rotation of the power piston is provided by a induction spin motor at a rotational speed of 30 Hz. A viscous coupling between the power piston and the displacer causes rotation of the displacer. The engine and the linear alternator are contained in a common pressure vessel to allow the conversion system to be hermetically sealed. Neodymium-Iron (28 megagauss) was selected as the permanent magnet material along with oriented silicon steel (M4) for the inner and outer laminations. An active cooling system is incorporated around the stator to assure the proper alternator temperature. A passive cooling system (fluid diodes) has been added to circulate the working gas (helium) around the alternator to insure cooling for the magnets and maintain the cold end of the engine below 80 °C. The linear alternator is connected to a series tuning capacitor and to the grid through an autotransformer. The single phase linear alternator has been designed to provide 26.2 kW<sub>e</sub> (nominal) and 31.5 kW<sub>e</sub> (peak) to the utility grid. A detailed manufacturing and costing analysis was completed, showing the free-piston Stirling conversion system could be manufactured using existing technology and meeting DOE's current cost goals.

#### Vibration Absorber

A passive (with no active feedback) vibration absorber was designed for the single piston free-piston Stirling conversion system. The balancer is a spring/mass system with the resonant frequency tuned to the 60 Hz engine operating frequency. The annular mass is coupled to the engine housing with multiple springs.

#### Cooling System

The heat rejection system is a closed loop system which include a radiator(s), a blower fan(s) and a coolant pump. The circuit is divided into two parallel loops: (1) for the engine cooler and (2) for the alternator stator and uses a single water pump. Based on the ASCS requirements and a trade study by McCord, a cooling system was designed utilizing industrial components which will maximize life and minimize maintenance requirements. The cooling system is dish mounted. A detailed evaluation of the cooling system resulted in a manufacturing cost of \$940 per unit in 1984 dollars.

#### Power Conditioning and Control System

The linear alternator is connected to the utility grid through an autotransformer and a series tuning capacitor. The frequency and autotransformer output voltage are essentially constant and established by the grid. The voltage at the alternator terminals is adjusted to match the changes in power while keeping the heater head temperature (700 °C) at its design point. The series tuning capacitors used to compensate for the internal inductance of the linear alternator assure stable operation of the Stirling conversion system. The controls are designed for fully automatic, unattended operation of the ASCS.

#### **STC ASCS PRELIMINARY DESIGN**

#### STC Preliminary Design Team

STC was responsible for project management, completing the preliminary design of the free-piston Stirling hydraulic concept and system integration. Westinghouse Electric Corporation, Pittsburgh, PA, is the manufacturing partner for STC.

The STC ASCS is shown in Fig. 5. STC teamed with Sanders Associates, Nashua, NH for the solar receiver and Thermacore, Inc., Lancaster, PA for the heat transport system. Westinghouse Electric Corp., Hazard Management Group, Pittsburgh, PA was responsible for an independent FMECA and FTA of the STC ASCS. Consultants used during the preliminary design process were: Saaski Technologies, Inc., Seattle, WA, as a heat transport consultant, Westinghouse Hanford Co, Hanford, WA for the high temperature materials expertise, Gedeon Associates, Athens, OH, for the thermodynamic simulation and General Engineering, Center Line, MI provided an update on the manufacturing and cost analysis for the STC design changes.

#### Receiver/Heat Transport System (HTS)

During the conceptual design phase of the ASCS Project, STC conducted an extensive trade study for the receiver/heat transport system and selected a reflux boiler. The reflux boiler appears to provide a simple and uniform method for transferring the heat to the Stirling engine (Ref. 11). Disadvantages cited for the pool boiler were the large liquid metal inventory along with the lack of demonstrated design theory. The major technical concern was the inability to predict stable boiling under all operational conditions for this unusual configuration. Subsequent subscale testing by Sandia showed that an enhanced surface could provide nucleation sites resulting in stable boiling (Ref. 12). Recent testing of a full scale boiler by Sandia has shown that stable boiling is achievable (Ref. 12).

The receiver/HTS is a pool boiler as shown in Fig. 6. The reflux boiler evaporator is designed to operate continuously and provides heat to the Stirling engine at 704 °C. A nominal insolation of 950

$W/m^2$  provides about 75 kW of thermal energy while a maximum peak insolation of  $1100 W/m^2$  provides about 87 kW of thermal energy to the absorber surface. The working fluid is a eutectic NaK alloy. Inconel 625 has been selected for the absorber and the body of the HTS. Nucleation sites are provided to the evaporator surface by adding a sintered metal wick. Thermacore results are discussed in detail in reference 13. At a production rate of 10 000 units per year the current pool boiler design has the potential to be produced at \$459 per unit in 1984 dollars.

#### Engine/Hydraulic System

The free-piston Stirling hydraulic engine (STIRLIC<sup>TM</sup>) converts heat delivered to the engine into high pressure hydraulic fluid. A cross section of the STIRLIC<sup>TM</sup> engine is shown in Fig. 7. The Stirling engine delivers high pressure hydraulic fluid to a commercial pump/motor which is coupled to a commercial induction generator to provide electrical energy to the utility grid. The heater head is designed to operate at 704 °C with an engine operating frequency at 50 Hz. The working fluid is Helium at a mean pressure of 18.3 MPa (2690 psi) which is hermetically sealed through the use of metal bellows seals used in the engine. The metal bellows seals are pressure balanced to ensure long life. The power pistons, the displacer rod and the stabilizer/controller are fully immersed in hydraulic fluid which provides full film lubrication of all sliding parts.

STC and their consultants conducted an extensive materials evaluation and trade study for the heater head assembly. Waspaloy was selected as the heater tube material and Inconel 713LC was selected as the heater head material (Ref. 14) for the prototype engine. Udimet 720 is an alternative material being considered for the heater head. Further, the STIRLIC<sup>TM</sup> engine conversion system underwent several value engineering improvements to enhance the manufacturability, simplify the design and reduce cost. The Stirling conversion system cost is estimated at \$2378 in 1984 dollars.

#### Hydraulic Pump/Induction Motor

Evaluation of two commercial variable displacement motors resulted in the selection of a high efficiency (93.1 percent) Volvo unit coupled to a commercial rotary induction generator. The three phase induction generator has been sized to provide up to 30.0 kW<sub>e</sub> (peak) to the utility grid.

#### Cooling System

Evaluation of the industrial components required for the cooling system resulted in the selection of a longer life and lower cost heat exchanger (radiator). The conventional (braze joints) radiator which is subject to corrosion from the glycol/water solution is replaced with a commercial radiator (all welded) from Modine Manufacturing Company.

#### Power Control System and Controls

Automatic regulation of the engine is accomplished via a integral control valve for the hydraulic pump which controls engine speed and balances the heat absorbed from the reflux boiler by the engine. Frequency and voltage are maintained by the utility grid. Power factor (PF) correction has been added to maximize power production at low power levels. Analysis has shown that the addition of 5 KVAR of capacitance to the power conditioning is cost effective for the ASCS and will maintain the PF above 0.85 during all regimes of operation. The ASCS control system will be designed to be fully automatic for unattended operation.

### CONCLUDING REMARKS

Both the Cummins and STC Preliminary Designs have met the ASCS Project design requirements. The Cummins ASCS features a heat pipe receiver integrated with a free-piston Stirling engine/linear alternator conversion system. The STC ASCS features a reflux boiler receiver integrated with a free-

piston Stirling/hydraulic conversion system. Cummins and Westinghouse have identified a niche market for dish Stirling systems for the worldwide remote electric power applications. Both organizations believe that successful operation in these niche markets may play a major role in the introduction of Stirling products into the commercial marketplace.

#### REFERENCES

1. Stein, William B.: Progress in Parabolic Dish Technology. SERI/SP-220-3237, 1989.
2. Dudenhoefer, J.E.: Programmatic Status of NASA'S CSTI High Capacity Power Stirling Space Power Converter Program. NASA TM-103142, 1990.
3. Walker, G.: Stirling Cycle Machines. Clarendon Press, Oxford, 1973, pp. 121-127.
4. Bean, J. R.: Development of the CPG 5 kW Dish/Stirling System. Proceedings of the 25th IECEC., Reno, NV., August 1990, Vol. 6, pp. 298-302.
5. Ross, B. A., et al: Commercialization of Dish-Stirling Solar Terrestrial Systems. Proceedings of the 25th IECEC., Reno, NV, August 1990, Vol. 5, pp. 292-296.
6. Shaltens, R. K. and Schreiber, J. G.: Preliminary Designs for 25 kWe Advanced Stirling Conversion Systems for Dish Electric Applications. NASA TM-103188, 1990.
7. Cummins Engine Company, Preliminary Design Review held at the NASA Lewis Research Center, Cleveland, Ohio. November 8 & 9, 1989.
8. Stirling Technology Company, Preliminary Design Review held at the NASA Lewis Research Center, Cleveland, Ohio. November 2 & 3, 1989.
9. Shaltens, R.K.: Advanced Stirling Conversion System for Terrestrial Applications. NASA TM-88897, 1987.
10. Shaltens, R.K. and Schreiber, J.G.: Comparison of Conceptual Designs for 25 kW<sub>e</sub> Advanced Stirling Conversion Systems for Dish Electric Applications. NASA TM-102085, 1989.
11. 25 kW<sub>e</sub> Solar Thermal Stirling Hydraulic System. DOE/NASA/0371-1, NASA CR-180889, 1988.
12. Diver, R. E., et al. Trends in Dish-Stirling Solar Receiver Designs. Proceedings of the 25th IECEC., Vol. 5, pp. 303-310, August 1990.
13. Anderson, W. G., et al.: Pool Boiler Heat Transport System for a 25 kWe Advanced Stirling Conversion System. Proceedings of the 25th IECEC., Vol. 5, pp. 268-273, August 1990.
14. Noble, J.G., et al.: Materials For A Stirling Engine Heater Head. Proceedings of the 25th IECEC., Vol. 5, pp. 281-291, August 1990.

**TABLE 1 - DOE COST<sup>a</sup> GOALS**

	Current Technology	Long-term Goals
Receiver, dollars/m <sup>2</sup>	70	40
Conversion, dollars/kWe	380	300
ASCS total, <sup>b</sup> dollars/kWe	646	452
Energy cost, dollars/kWe	0.13	0.05

**a** 1984 Dollars.

**b** Receiver (11-m concentrator) and conversion (nominal 25 kWe) combined.

**TABLE II - COMPARISON OF ASCS PRELIMINARY DESIGNS**

	CEC	STC
Heat Supplied (peak), kWt	86.8	86.8
Electrical power (peak), kWe	31.5	25.9
Annual Energy, kWh	65 953	57 100
Annual Efficiency, percent	31.8	27.6
Receiver	Heat Pipe	Reflux Boiler
Receiver efficiency, percent	92.0	89.6
Liquid metal	Sodium (Na)	NaK
Heater head temperature, °C (K)	700 (973)	704 (977)
Cooler temperature, °C (K)	60 (333)	54 (327)
Temperature ratio, Th/Tc	2.92	2.99
Efficiency (solar to electric), percent	34.9	29.8
Engine frequency, Hz	60	50
Pressure, MPa (psi)	10.5 (1540)	18.3 (2690)
Working fluid	Helium	Helium
Power conversion	Linear Alternator	Hydraulic pump w/Induction Gen.
Electrical efficiency, percent	94.0	95.0
Electrical output	240 V,1 $\phi$	240 V,3 $\phi$
Power conditioning	Autotransformer	5 KVAR Capacitance
Controls	Automatic	Automatic
Weight on TBC, kg (lb)	1791 (1740)	864 (1900)
ASCS total cost, 1984 dollars	Under review	6931



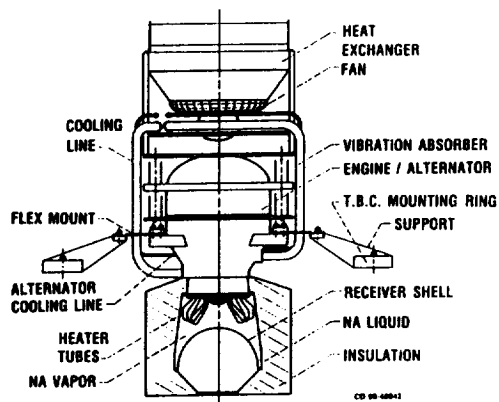


FIGURE 2. - CUMMINS' SINGLE PISTON STIRLING ENGINE ASCS LAYOUT.

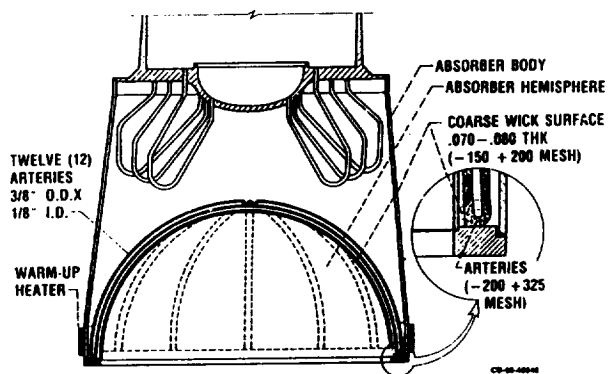


FIGURE 3. - HEAT PIPE RECEIVER/HEAT TRANSPORT SYSTEM.

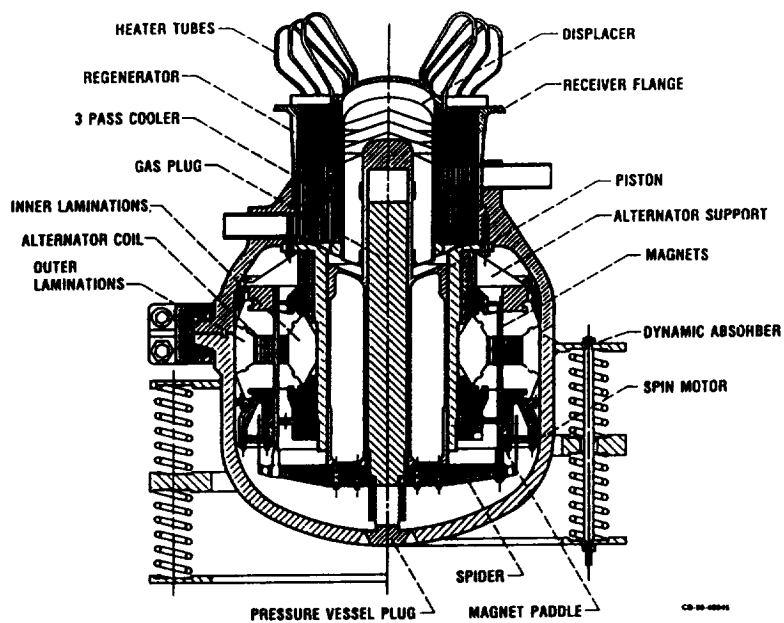


FIGURE 4. - CUMMINS' FREE-PISTON STIRLING/LINEAR ALTERNATOR DETAILS.

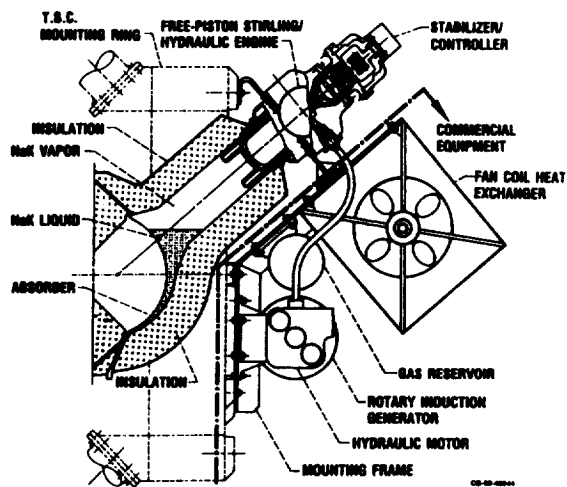


FIGURE 5. - STC'S STIRLING/HYDRAULIC ASCS LAYOUT.

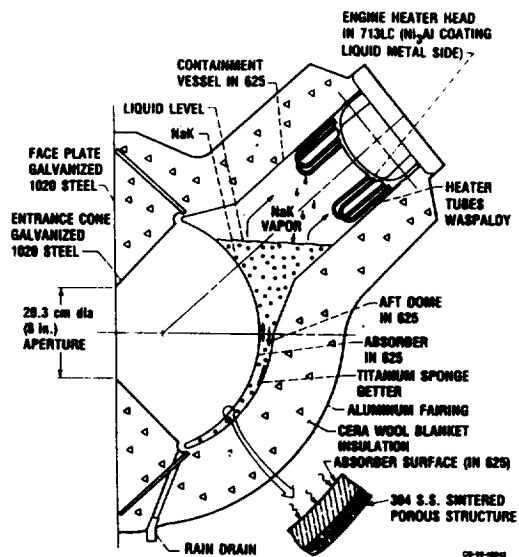


FIGURE 6. - REFLUX BOILER RECEIVER/HEAT TRANSPORT SYSTEM.

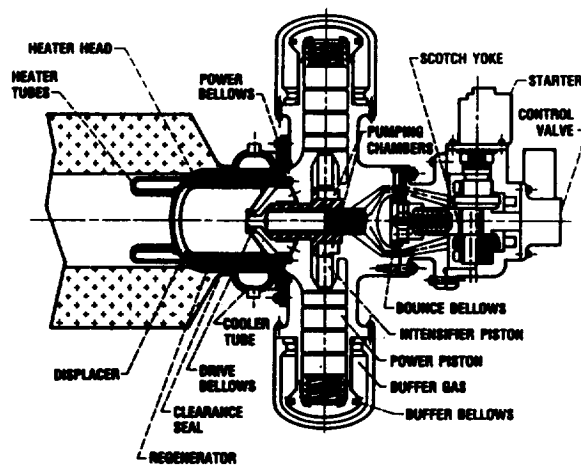


FIGURE 7. - STC'S FREE-PISTON STIRLING/HYDRAULIC DETAILS.

## FOUR QUADRANT CONTROL OF INDUCTION MOTORS

Irving G. Hansen  
National Aeronautics and Space Administration  
Lewis Research Center  
Cleveland, Ohio 44135

## ABSTRACT

Induction motors are our nation's workhorse, being the motor of choice in most applications due to their simple rugged construction. It has been estimated that 14 to 27 percent of the country's total electricity use could be saved with adjustable speed drives. Until now though, induction motors have not been well suited for variable speed or servo drives, due to the inherent complexity, size and inefficiency of their variable speed controls. Work at NASA Lewis Research Center on field oriented control of induction motors using a pulse population modulation method holds the promise for the desired drive electronics. The system allows for a variable voltage to frequency ratio which enables the user to operate the motor at maximum efficiency, while having independent control of both the speed and torque of an induction motor in all four quadrants of the speed torque map. Multiple horsepower machine drives have been demonstrated, and work is on-going to develop a 20 hp average, 40 hp peak class of machine. This paper discusses the pulse population technique, results to date, and projections for implementation of this exciting new motor control technology.

## BACKGROUND

For aerospace and launch vehicles, electric actuators have fundamental advantages over the more conventional hydraulic actuators. Basically they are more efficient, and require less maintenance. Until recently technical limitations have denied their aerospace application at the required high power levels. Now, however, new technology is available which allows 50 kW class electric powered flight actuators to be built today. At NASA Lewis Research Center electric actuator work for aerospace applications has concentrated on induction motors and associated high frequency link drives, which should have dramatic payoffs in any mechanical process requiring variable operating points.

## WHY INDUCTION MOTORS?

The induction motor is, by a wide margin, the most commonly applied motor. It is a mechanically simple, rugged machine capable of providing rapid response and high peak torques. In its simplest conceptual form it is essentially a multiphase transformer with its secondary (rotor) shorted and free to move axially with the rotating magnetic field (stator flux) figure 1.

In such a motor the magnetic flux of the stator varies directly with the source voltage and inversely with the source frequency. The current in the rotor results from the voltage induced by the stator flux and the rotational speed difference between the stator flux field and the rotor (slip frequency). As usually encountered the motor is operated from a fixed voltage, fixed frequency source (e.g., 120/208 V and 60 Hz). When operated under these conditions, a typical induction motor displays the characteristics shown in figure 2. Of particular importance is that under these operating conditions, high torque, efficiency, and power factor are realisable over only a rather limited range of speed.

Of interest also is that these characteristics are symmetrical about the stator (synchronous) speed, and that, for example, if the rotor is driven above synchronous speed (fig. 3) generator

action results. The optimum operating region discussed earlier is expanded in figure 4 with some of the data replotted. The ratio of torque to current is a very interesting parameter. An actuator motor is a current to torque transducer being driven by power electronics with sharply defined current limits. Therefore, in order to fully utilise the electronics capabilities achieving the maximum torque/amp is crucial. Also important in aerospace applications is the minimisation of losses to enhance thermal control. To better illustrate this point under constant voltage and frequency operation refer to figure 2. Full rated torque occurs at about 0.75 power factor and an efficiency of about 0.5. However, maximum torque/amp occurs at about 0.8 power factor and an efficiency of about 0.75. It is to be noted that for this set of data minimum loss is different than maximum efficiency, which is meaningless at stall (blocked rotor) speed.

### WHAT TO DO ABOUT IT?

There are several ways to relocate the optimum operating point of an induction motor. If the frequency is held constant and the voltage varied, the characteristics of figure 5 are obtained. This is the approach taken by Frank Nola (ref. 1) to improve the power factor and efficiency of lightly loaded motors. This provides essentially variable torque, constant speed operation. If both the voltage and the frequency are varied with their ratio held constant  $v/f = c$  (constant stator flux), the characteristics of figure 6 are obtained. This represents variable speed operation with constant peak torque.

Combining both strategies provides the capability of optimizing the motor characteristics at any speed and load including blocked rotor operation in an actuator. The capability of optimising operation at full torque and low speed is a very important advantage. Since  $v/f$  control allows full torque at zero shaft speed, the rotor resistance and its associated loss may be reduced. Figure 7 shows how high rotor resistance usually required for high starting torque may be greatly reduced. If  $v/f$  control is implemented, the motor peak torque may be obtained at zero speed and rated current. This typically reduces starting currents by a factor of six.

### WHAT'S THE PROBLEM?

The approach generally taken to provide a variable frequency source is to create time phased square wave voltages, which when applied line to line form a quasi-sine wave (six-step) (fig. 8). Such an approach, however, has serious limitations. The voltage is not sinusoidal and its harmonic distortion causes additional losses in the motor. Also, the semiconductor switches must turn off the motor current and operate under high stress. Then, to achieve voltage control the voltage steps are pulse width modulated (PWM), which increases both the stress and loss (fig. 9).

Aerospace actuators now in service generally use permanent magnet (PM) motors (brushless dc). No capability exists to optimise either efficiency or power factor. In these motors the resulting high currents, which produce loss but no output, even further stress the switching elements. One approach to these problems requires energy controlling snubbers across the switches. If this snubber energy is recovered, high current spikes are created that add to the already severe electrical noise problem.

Since these problems, in general, involve inductive energy, which varies with the square of the current, they rapidly worsen with increasing power levels. As a result, today's conventional PWM-PM motor technology scales poorly; and, actuators have been limited to relatively low power applications. An attempt to increase power level results in electronic packages many times larger than the associated motor.

## WHAT'S NEW?

NASA Lewis, together with its contractors, has developed high power high frequency power distribution systems and components (refs. 2 and 3). Conceptually, such a system delivers energy tailored to the user's actual requirements with the least number of conversions possible (fig. 10). Of particular importance is the fact that in a typical 20 kHz system energy flow is controlled and quantised into 25  $\mu$ s half sine pulses. In a 10 kW controller each half sine pulse represents only 1 J of energy. Using a high frequency sinusoidal power source, lower frequencies (including dc) may be synthesised. All switching functions are performed at zero voltage, which greatly reduces the circuitry and the electrical loss (fig. 11). The synthesis scheme shown is referred to as pulse population modulation. In this scheme the voltage is controlled by the density of the pulse population and the frequency is controlled by the pulse pattern.

With this modulation scheme voltage and frequency may be independently varied, which in turn allows independent control of the rotor current and the stator current (refs. 5 and 6). Actual bidirectional, four quadrant operation of a 20 hp induction motor is shown (fig. 12). This particular motor controller combination is presently under test at General Dynamic Space Systems as part of the Advanced Launch Development Program (ALDP). This photograph was taken while a large inertia load was being driven bidirectionally under constant torque. The step is the torque command and the ramp is the resulting shaft speed. Of particular note is the instantaneous torque response and the torque control at zero speed. Work has started on a 40 hp electrical actuator using this technology. An advanced induction motor is presently being designed for the actuation system.

## WHO CAN USE IT?

The combined benefits of low loss, rapid response, and high power factor, together with electronic switch protection, makes this technology a candidate for many motor driven processes. Immediate applications identified are for thrust vector control on the ALS and other missile systems.

Efforts are now under way at Douglas Aircraft to evaluate the impact of this technology, not only to flight actuators, but also to many other aircraft loads, such as nose wheel steering, braking, and environmental control. This is a renewal of an all electric airplane effort begun at NASA Lewis in the mid-1980's (ref. 7).

## CONCLUSIONS

This technology is rapidly approaching full maturity. It has a sound analytical base, is supported by component developments, and is being verified by full scale demonstrations.

## REFERENCES

1. Nola, F.J.: Power Factor Control System for AC Induction Motors, U.S. Patent 4266177, May 1981.
2. Hansen, I.G., and Wolff, F.J.: 20 Kilohertz Space Station Power System. NASA TM-88801, 1986.
3. Hansen, I.G.: Status of 20 kHz Space Station Power Distribution Technology. NASA TM-100781, 1988.

4. Sood, P.K.; Lipo, T.A.; and Hansen, I.G.: A Versatile Power Converter for High-Frequency Link Systems. *IEEE Trans. Power Electron.*, vol. 3, no. 4, Oct. 1988, pp. 383-390.
5. Roth, M.E.: Electromechanical Actuation for Thrust Vector Control Applications. NASA TM-102548, 1990.
6. Burrows, L.M.; Zinger, D.S.; and Roth, M.E.: Field Oriented Control of Induction Motors. NASA TM-103154, 1990.
7. Hoffman, A.C., et al.: Advanced Secondary Power System for Transport Aircraft. NASA TP-2463, 1985.

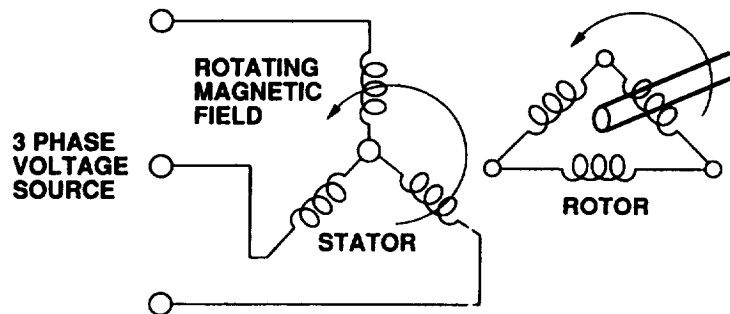


FIGURE 1. - INDUCTION MOTOR.

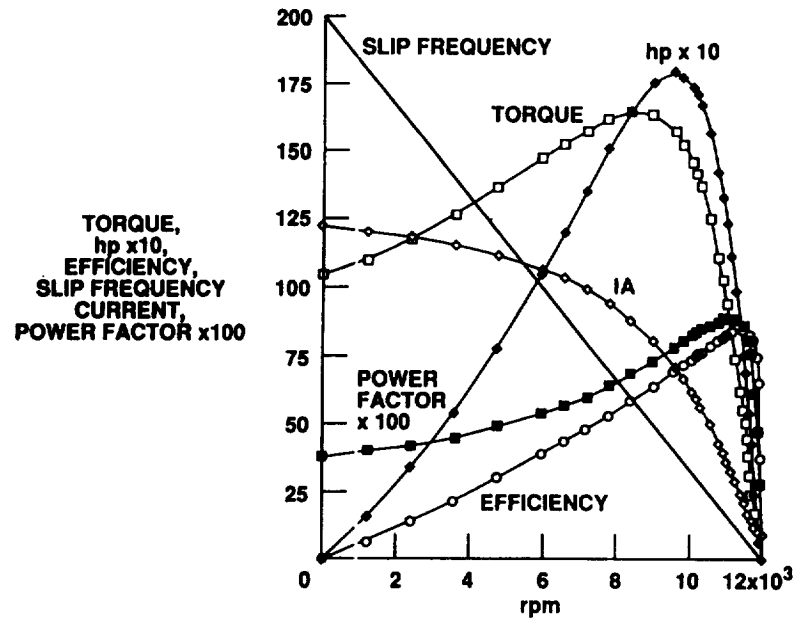


FIGURE 2. - INDUCTION MOTOR CHARACTERISTICS, FIXED VOLTAGE AND FREQUENCY.

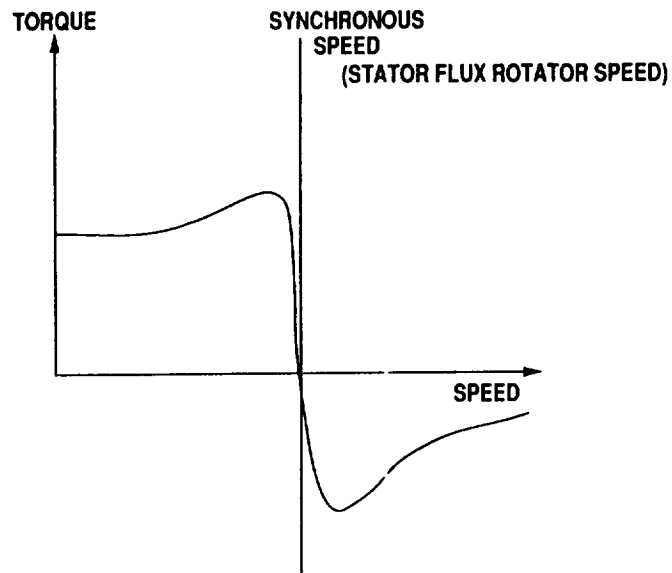


FIGURE 3. - INDUCTION MOTOR/GENERATOR CHARACTERISTICS.

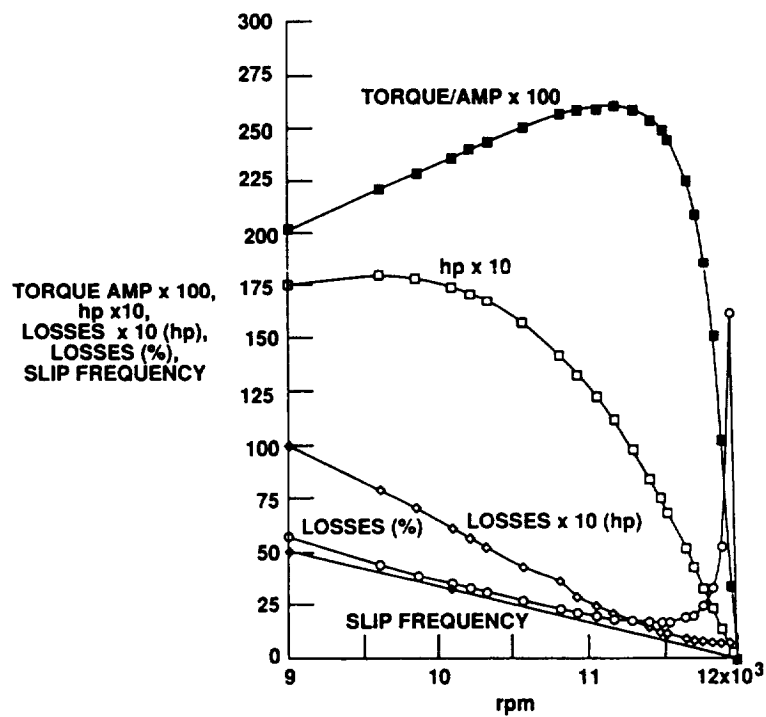


FIGURE 4. - DATA FOR MOTOR 939D228-4.

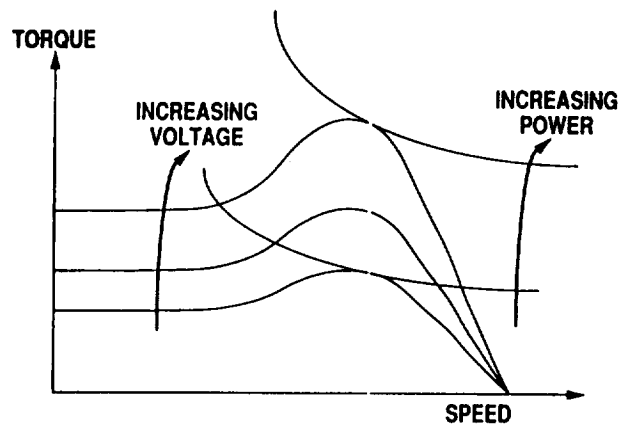


FIGURE 5. - VARIABLE VOLTAGE CHARACTERISTICS.

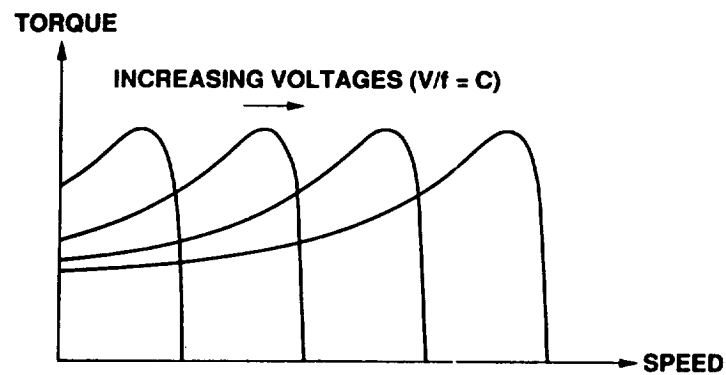


FIGURE 6. - CONSTANT  $V/f$  INDUCTION MOTOR.

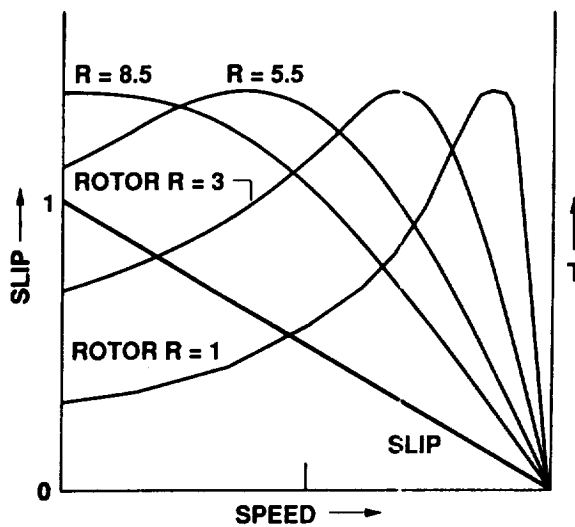


FIGURE 7. - SPEED TORQUE CHARACTERISTICS OF INDUCTION MOTOR AS FUNCTION OF ROTOR RESISTANCE.



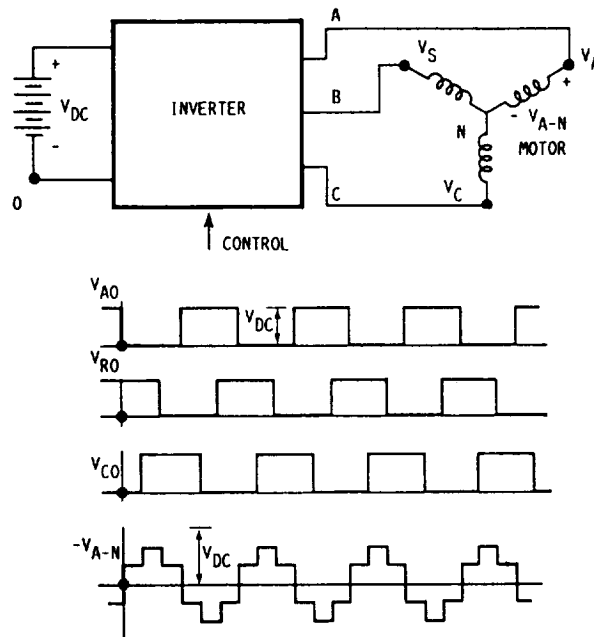


FIGURE 8. - GENERATION OF SIX-STEP SQUARE WAVE.

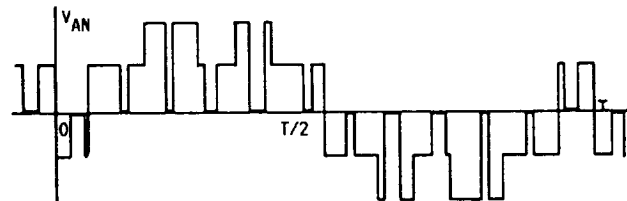


FIGURE 9. - GENERATION OF SINE WAVE PWM.

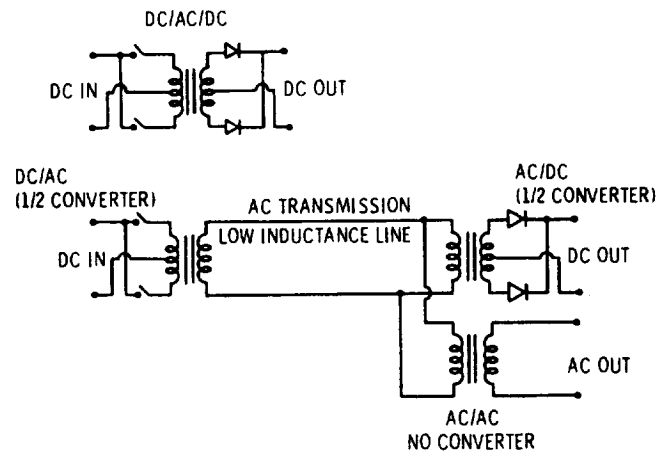
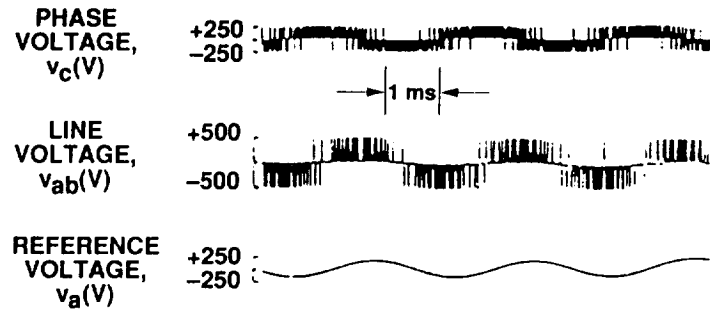


FIGURE 10. - "SPLIT" SYSTEM CONCEPT TO CHANGE DC VOLTAGE LEVELS (DC/DC CONVERTER).



COMPUTER SIMULATION OF LOW FREQUENCY SYNTHESIS  
(UNIVERSITY OF WISCONSIN)

FIGURE 11. - PULSE POPULATION MODULATION, PPM.

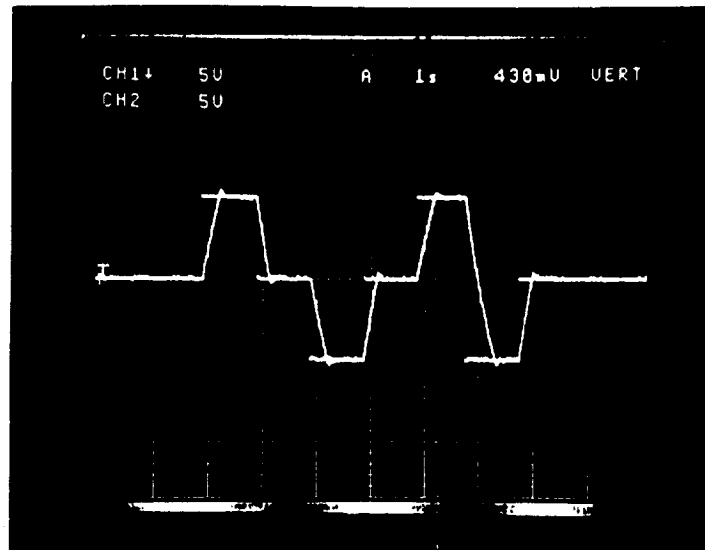


FIGURE 12. - FOUR QUADRANT - CONSTANT TORQUE OPERATION AT 20 HORSEPOWER.

ORIGINAL PAGE IS  
OF POOR QUALITY

- ADVANCED LAUNCH SYSTEM ACTUATORS
- SHUTTLE ACTUATORS
- AIRCRAFT: POWER BY WIRE  
NOSEWHEEL STEERING  
STARTER GENERATORS  
AUXILIARY POWER UNIT
- SPACE POWER SYSTEMS AND INTERFACES  
WITH BRAYTON AND STIRLING ENGINES
- OTHER APPLICATIONS, ANY MECHANICAL  
PROCESS REQUIRING A VARIABLE THROUGHPUT

FIGURE 13. - ONGOING EFFORTS.

APPLYING HIGH FREQUENCY POWER CONTROL  
TECHNIQUES TO INDUCTION MOTOR CONTROL  
REPRESENTS ENABLING TECHNOLOGY DISPLAYING THE  
ADVANTAGES OF:

- LOW NOISE, BOTH ELECTRICALLY AND ACOUSTICALLY
- LOW WEIGHT
- LOW VOLUME
- HIGH EFFICIENCY
- HIGH POWER LEVEL

FIGURE 14. - CONCLUSIONS.

**HIGH POWER BIPOLAR LEAD-ACID BATTERIES**

**Gerald Halpert and Alan Attia  
Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, CA 91109**

**INTRODUCTION**

The Jet Propulsion Laboratory (JPL), with interest in advanced energy storage systems, has been involved with the development of a unique lead acid battery design. This battery utilizes the same combination of lead and lead dioxide active materials present in the automobile starting battery. However, it can provide 2-10 times the power while minimizing volume and weight. The typical starting battery is described as a monopolar type using one current collector for each plate. The bipolar battery uses a single current collector for both the positive and negative plate of adjacent cells. Specific power as high as 2.5 kW/kg has been projected for 30 second periods with as many as 2000 recharge cycles.

**BACKGROUND**

In a 1985 study prepared for the Department of Energy (1), 34 alternative five-passenger advanced vehicles were ranked using multi-attribute decision analysis. The ranking, designed to ascertain the most promising battery EV technology/range combinations, indicated that the bipolar lead-acid 100-mile-basis of the ranking was the exceptionally high power capability, excellent thermal characteristics, well known chemistry, low cost, large industrial base and potential to develop the technology. A number of patents have been issued on various aspects of the cell and battery technology.

In July of 1986, a program was initiated at JPL under the sponsorship of the Air Force Wright Aeronautical Laboratories for the development of a high specific power, 50-kW, sealed bipolar lead-acid batter. This four-year program was the result of the A.F. interest in the JPL bipolar lead-acid technology developed during the mid 1980s. In the fall of 1990, three 50 kW modules of a quasi-bipolar design (a modification of the true bipolar battery), based on the JPL technology, were assembled under a subcontract to Johnson Controls Incorporated. At this writing, one was delivered to the A.F. for testing.

**DESCRIPTION OF SYSTEM**

The unique aspects of this technology are the sealed bipolar construction and the light weight composite biplates utilizing electrically conductive tin dioxide coated glass fibers.

The basic building block (Figure 1) of the bipolar design is a light-weight composite bipolar plate (biplate) consisting of two layers (substrates). One layer is composed of conductive tin oxide-coated glass fibers and the other of carbon fibers, each embedded in its own polymeric matrix. The negative substrate consists of a conductive carbon fiber or powder filled impervious thermoplastic composite. It serves as the current collector (substrate) for the negative plate (Pb). The carbon-based materials are commercially available. The unique positive substrate serving as the current collector for the positive electrode (PbO<sub>2</sub>) consists of tin dioxide coated glass fibers (Figure 2) or powders imbedded in a non-porous thermoplastic and molded into a composite to provide the necessary conductivity. At the present time, commercial sources of tin dioxide coated fibers are not available. The substrates must be capable of allowing electrons resulting from the discharge reactions to transfer from the negative active materials of one cell to the positive active materials in the adjacent cell (and in reverse during charge).

A cell (Figure 3) consists of the positive side of one biplate facing the negative side of the adjacent biplate. Between the active materials in each cell is a porous glass mat separator which serves as a reservoir for the electrolyte and to physically separate the positive and negative active materials. The biplates

assembled in series are sealed at their periphery and together with the current collector endplates produce the sealed battery. The design is such that the current travels from one end of the battery directly through the biplates to the negative end. There are no tabs or cell terminals which minimizes the resistance to current flow.

## DESIGN CONSIDERATIONS

The requirement of maximum specific power dictates a bipolar construction in order to eliminate resistance through the lead grid of conventional lead-acid batteries. For any application, either terrestrial or space, such a battery also needs to be sealed, as watering would be impractical because of the large number of very thin cells required for a multi-kW battery. Lead has been the commonly used substrate for preparation of the bipolar plates in previous lead-acid batteries, but the weight of lead is unacceptable in minimum practical thicknesses.

Currently, the conductive glass fibers are produced by spraying a solution of tin tetrachloride in the presence of air on a heated glass fiber mat (500°C). This process, known as spray pyrolysis, has long been a production method for applying transparent tin dioxide (SnO<sub>2</sub>) conductive thin films to glass. In this instance, glass fibers of 10 micron diameter are coated with 0.5 to 1 micron of conductive tin oxide. The material has been shown to be thermodynamically stable with respect to oxidation or reduction at the positive electrode in normal operation. The material will never be life-limiting when used in the positive plate environment if care is taken to prevent reversal of the positive plates. The new light-weight plastic composites developed by JPL have an overall thickness of 0.06 cm. The goals of the volume resistivity of 1 to 2 ohm-cm and an area density of 0.15 g/cm<sup>2</sup> have been achieved.

The capability for sealing is provided by the gas recombination mechanism associated with the "starved" cell design. That is, unlike the typical automotive starting battery, there is no free flowing electrolyte. All of the sulfuric acid electrolyte is contained in the pores of the active materials and separator allowing for gas to freely circulate in each cell. The sealing can be accomplished because the oxygen gas generated on over charge at the positive electrode is recombined at the negative electrode in the same cell, thus, minimizing gas pressure. The design requires the appropriate balance between the capacities of the positive and negative plate in each cell.

## ADVANTAGES

The fundamental advantage of the bipolar lead-acid couple for high power applications is due to four characteristic: high open circuit voltage, low electrolyte resistivity, very low cell-to-cell resistance, and discharge with an increase in entropy. Very few couples can match the power capability of the lead-acid system (given by the ratio of the square of the open circuit voltage to the electrolyte resistivity) and most couples discharge exothermically. Because of the increase in the entropy of the system during discharge, the lead-acid battery will absorb heat during discharge. Furthermore, the sealed bipolar lead-acid battery described above has the advantages of very light weight construction, a thermodynamically stable conductor in the bipolar plate, and the absence of auxiliary equipment such as pumps, cooling loops, and heavy storage containers required by several other battery systems.

## FUNDAMENTAL CONCEPTS

The important value of specific power (kW/kg) of a high power battery can be factored into two parts: the surface power density (W/cm<sup>2</sup>) and the specific area (cm<sup>2</sup>/kg). Batteries very often perform well in one factor but generally do not fare well in both characteristics. For comparison purposes, the maximum surface power density for an electrochemical cell can be expressed as a figure of merit, M, as given by:

$$M = \frac{(E'_{OC})^2}{\rho} \quad (1)$$

where  $E'_{oc}$  is the battery voltage extrapolated to zero current from the polarization curve, and  $\rho$  is the electrolyte resistivity. For comparison,  $M$  is calculated for nine different couples and shown in Table 1. Based on intrinsic properties of the couple alone the lead-acid system has three times the surface power density of the Ag-Zn and four times the Ni-H<sub>2</sub> and Ni-Cd system.

For maximum specific power the area factor must be a maximum. This can only be accomplished by minimizing the thickness and weight of the components. This is done by use of a light weight bipolar plate and the use of thin layers of active material for the electrodes.

Table 1 also shows the calculated value for the quantity  $Q/E$  which is the ratio of thermal ( $Q$ ) to electrical ( $E$ ) energy when discharging at  $2/3 E'_{oc}$ . The values for  $Q/E$  ratio are consistent with those projected from thermodynamic considerations and show a clear superiority of the lead-acid system, i.e., there is a lower heat generation for the electrical energy generated.

### PERFORMANCE PROJECTIONS

The mean specific power (MSP) of the battery can be represented by:

$$MSP = \frac{E_d i}{M} \quad (2)$$

where  $E_d$  is the discharge voltage,  $i$  is the current density and  $M$  is the total battery mass per unit of plate area. The current density is related to the discharge time ( $t_d$ ) by:

$$i = \frac{0.2241 m^+ u^+}{t_d} \quad (3)$$

The factor 0.2241 is the Ah/g of positive active material (limiting electrode), and  $m^+$  and  $u^+$  are the mass/area (g/cm<sup>2</sup>) and utilization efficiency of the positive active material (PAM) respectively.

Computer generated plots of the relationship between MSP and positive plate thickness are given in Figure 4 for various pulse lengths. The projections also include the effect varying the mass/area and resistance of the positive active material. These curves show that the required pulse length will dictate the thickness of the positive active material and that the mass reduction has a greater effect than the resistance.

### STATUS OF HARDWARE DEVELOPMENT EFFORT

The designs for the bipolar plate and the battery were under development at JPL from 1986 to 1988, culminating with the construction of a small 6-cell bipolar battery and the fabrication of several versions of a light-weight composite bipolar substrate. In August, 1988, the responsibility for the A.F. effort of scaling up the design of the battery was assigned to Johnson Controls, Inc. (JCI), in Milwaukee. Under JPL's management, all critical areas in the development of the battery were addressed by JCI in three parallel tasks.

At the component level, the key critical area of development of a light weight bipolar substrate and frame stable in the environment of the lead-acid battery received the most emphasis. JCI developed a 2-layer substrate consisting of a thin carbon filled polyethylene layer for the negative side, bonded to a thin conductive tin oxide coated glass fibers filled plastic layer for the positive side. The use of thin coatings of lead to enhance the adhesion of the active materials to the conductive plastics was also investigated.

At the cell level, several critical aspects such as active materials formulation, separator selection, thin electrodes pasting and formation, sealed construction and gas recombination, were tested for refinements and improvements. At the battery level, critical factors such as module sizing and voltage, thermal management, battery configuration, termination, current collection and light weight end block support, were

investigated experimentally and were also explored using JCI's comprehensive computer model for the lead-acid battery. The battery design specifications are given in Table 2. The drawings and a photograph are given in Figures 5 and 6.

A modified 50-kW bipolar battery was delivered to the Air Force in September 1990 and is currently under test. Some compromises were made to meet the delivery schedule. The delivered design was a package containing 4-20 cell batteries in series. One major deficiency was in the biplate design, i.e., the biplate was a quasi-bipolar (folded screen) design and contained no stannic oxide fibers. Also, the plate area was reduced by half and the thickness of the active material increased for simplified assembly. This provided the required capacity but reduced the power capability.

Despite these significant deficiencies, the initial results showed considerable promise: the battery demonstrated a specific power in excess of 1 kW/kg for short discharges on the order of a few seconds, or more than twice the specific power achieved to date with conventional high power lead-acid batteries.

### ELECTRIC VEHICLE CONSIDERATIONS

The electric vehicle battery must satisfy numerous requirements including high specific power and energy, high reliability, low cost and maintainability and long cycle life at high depths of discharge. To these ends, the sealed bipolar lead-acid battery containing the stannic oxide material offers some new alternatives for potential use in the EV arena. Higher values of specific power and specific energy have been demonstrated when comparing with the lead-acid starting, lighting, ignition (SLI) battery. Simplicity of manufacturing this sealed battery offers comparable if not improvements in cost, reliability and maintainability. One other attribute is the endothermic property of the lead-acid system when it is discharged. This characteristic will result in less complexity of battery design compared to other batteries that are exothermic on discharge.

The subject of long cycle life at high depths of discharge must be addressed. This critical area is a function of several cell design factors, most notably the active material formulation. Most manufacturers tend to use their stock or patented materials; however, innovation will be required to design for this specific application. Substantial improvements in performance were obtained in preliminary experiments by the JPL Battery System Group with the addition of small amounts of tin oxide to the positive active mass.

There are two alternatives for using this battery in an EV application. One is the use of this battery as the main power source. Tradeoffs in design are required to balance all the factors. For acceleration the present high power design would be ideal; for distance, high capacity is required. Designing for both is possible with appropriate compromises. The second alternative involves the hybrid concept. The power system would utilize the Sealed Bipolar Lead-Acid Battery for acceleration together with a second power source for extended range. The second power source could be a deep discharge lead-acid battery, a fuel cell battery or even a small internal combustion (IC) engine.

Several studies have concluded that both EV alternatives including those with the range extenders are feasible. While improvements in these systems are required, with a properly designed sealed true-bipolar plate lead-acid battery containing the stannic oxide additive and the correct thickness/area balance and formulation of active materials, one of the major hurdles for an EV battery may have been cleared.

### REFERENCES

1. "Advanced Vehicle Systems Assessment", JPL Publication 84-79, prepared by the Jet Propulsion Laboratory, California Institute of Technology, for the Department of Energy, (1984).
2. U.S. Patents, 4,098,967 (1978); 4,275,130 (1981); 4,353,969 (1982); 4,405,697 (1983); 4,507,372 (1985); 4,510,219 (1985); 4,539,268 (1985); 4,542,082 (1985); 4,547,443 (1985); 4,625,395 (1986); and 4,658,499 (1987).

- CONDUCTIVE ( $< 2 \text{ OHM-CM}$ )
- LIGHTWEIGHT ( $< 150 \text{ MG/CM}^3$ )
- THIN ( $0.025''$  ( $0.0635 \text{ CM}$ ))
- CONDUCTIVE TO ELECTRODE ADHESION
- PORE FREE
- ELECTROCHEMICALLY STABLE AT  $\text{PbO}_2$  POTENTIAL
- LOW SOLUBILITY IN SULFURIC ACID
- HIGH OXYGEN OVERPOTENTIAL
- LOW COST

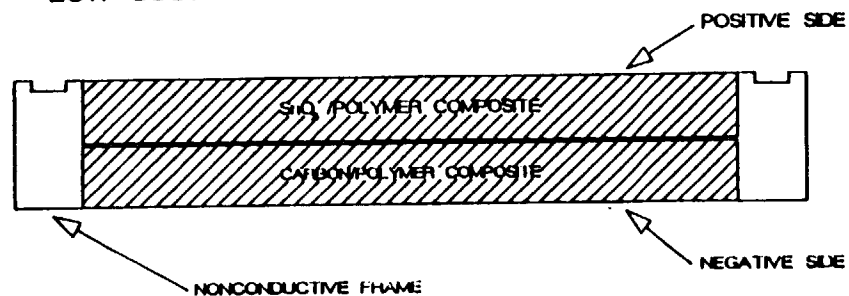
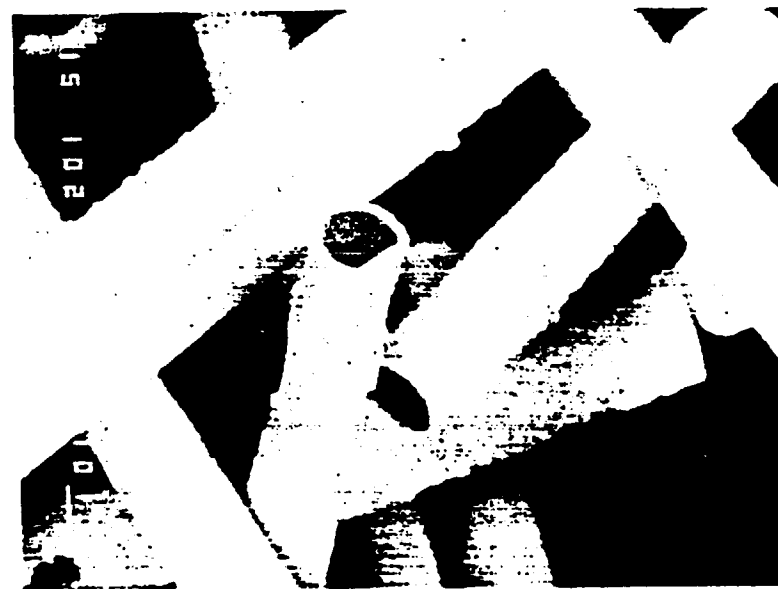


Figure 1. Bipolar Plate Design Goals



ORIGINAL PAGE IS  
OF POOR QUALITY

Figure 2.  $\text{SnO}_2$  Coated Glass Fibers

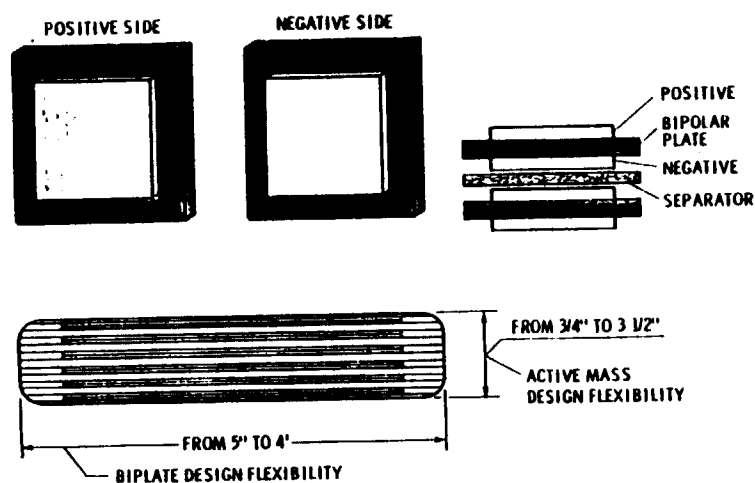


Figure 3. Bipolar Electrode Configuration

COUPLE	$E'_{oc}$ (VOLT)	$\rho$ (OHM-CM)	$M^{***}$ (WATTS/CM)	$Q/E$ (@ $2/3 E'_{oc}$ )
LEAD-ACID	2.10	1.12 (1)	3.93	0.46
LI (A1)-FeS	1.33**	0.64 (2)	2.76	0.65
H2 O2	1.05*	0.60 (3)	2.20	1.11
Ni-Zn	1.72*	1.84 (4)	1.61	0.54
Ag-Zn	1.69*	1.84	1.37	0.76
Ni-H2	1.32*	1.84	0.95	0.80
Ni-Cd	1.30*	1.84	0.92	0.58
Na-S	2.08**	4.7 (5)	0.92	0.55
LISOC12	3.32	47.6 (6)	0.28	0.68

$Q/E$  IS THE RATIO OF THERMAL TO ELECTRICAL POWER AT  $2/3 E'_{oc}$

- (1) VALUE AT 30 C
- (2) VALUE AT 450 C
- (3) VALUE AT 100 C
- (4) VALUE AT 18 C
- (5) VALUE AT 300 C
- (6) VALUE AT 60 C

\*EXTRAPOLATION FROM  $E - I$  CURVES

\*\*OPEN CIRCUIT VOLTAGE USED BECAUSE OF LOW ACTIVATION POLARIZATION

\*\*\* $M$  = FIGURE OF MERIT =  $(E'_{oc})^2/\rho$

Table 1. Characteristics of Electrochemical Couples

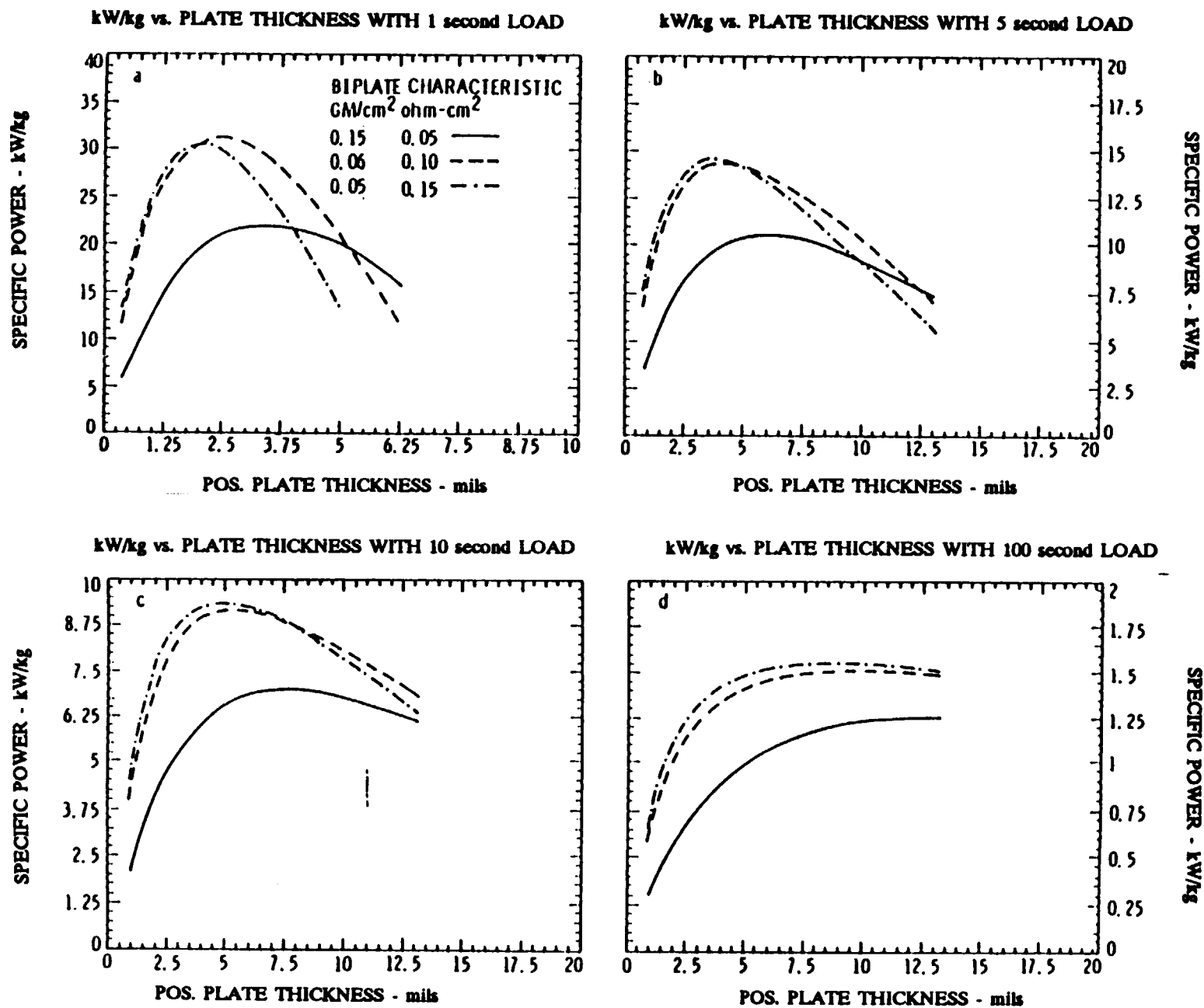


Figure 4. Specific Power vs. Positive Plate Thickness and Pulse Length



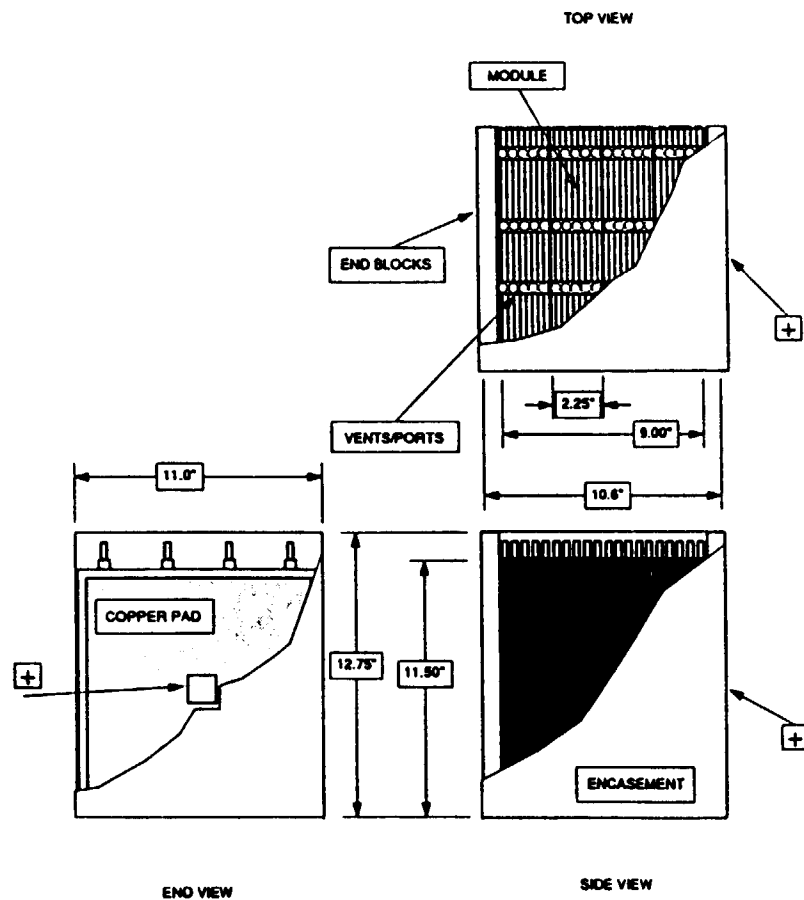


Figure 5. JPL/JCI/Air Force 50KW Module

Number of cells	80
Dimensions	10.6"x12.8"x11"
Weight	43.1 Kg
Substrate thickness	0.020"
Frame thickness	0.098"
Cell thickness	0.102"
Active area	521 cm <sup>2</sup>
Grid open area	82%
Grid thickness	0.020"
Acid density	1.305
Acid amount/cell	80 ml

Table 2. Battery Design Specifications

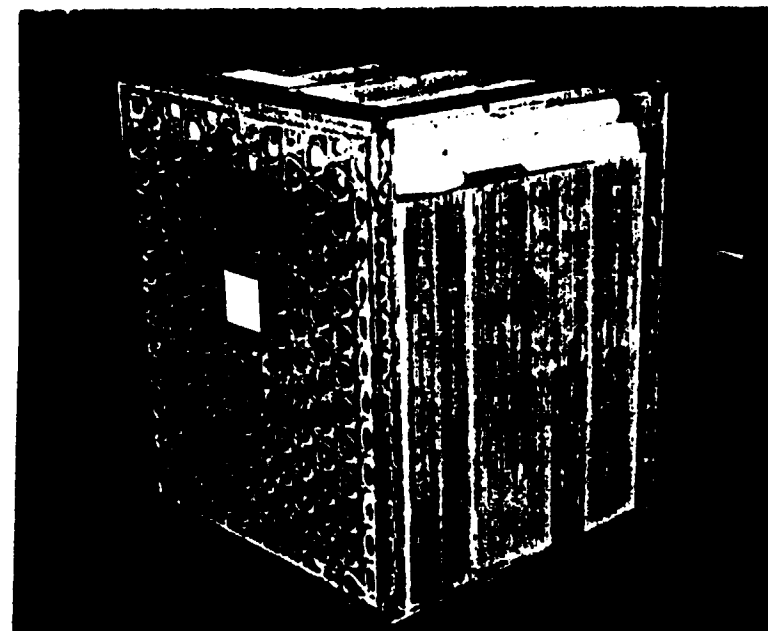


Figure 6. Photograph of the 50 KW Module

ORIGINAL PAGE IS  
OF POOR QUALITY

**Scientific CCD Technology at JPL**

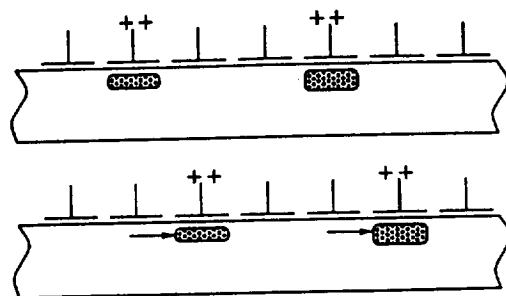
J. Janesick, S.A. Collins, and E.R. Fossum  
Imaging Systems Section  
Jet Propulsion Laboratory  
Pasadena, CA 91109

**Introduction**

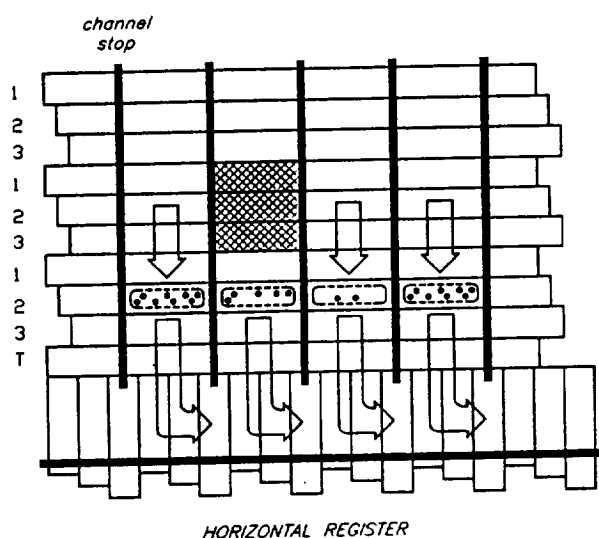
Charge-coupled devices (CCDs) were recognized for their potential as an imaging technology almost immediately following their conception in 1970. Twenty years later, they are firmly established as the technology of choice for visible imaging. While consumer applications of CCDs, especially the emerging home video camera market, dominate manufacturing activity, the scientific market for CCD imagers has become significant. This paper describes activity of JPL and its industrial partners in the area of CCD imagers for space scientific instruments. Requirements for scientific imagers are significantly different from those needed for home video cameras, and are described below. An imager for an instrument on the CRAF/Cassini mission is described in detail to highlight achieved levels of performance.

A charge-coupled device can be thought of as an electronic conveyor belt, that collects electrons generated by incident light, and shifts them toward an output amplifier. The charge is confined by voltages applied to electrodes separated from the semiconductor by a thin insulating layer. As the voltages on adjacent electrodes are changed, the charge in the semiconductor is pulled along (See Fig. 1). By using hundreds or thousands of electrodes, a packet of a few thousand electrons can be transferred over a centimeter or more of distance. A major issue in CCDs is the charge transfer efficiency (CTE) which is defined as the fraction of charge successfully transferred from electrode to electrode (though sometimes defined per pixel). While a CTE of 0.999 might appear to be rather good (losing only 1 electron in a thousand), if the packet undergoes 500 transfers, the net efficiency is  $0.999^{500} \approx 0.61$  representing an unacceptable deterioration of the signal. Scientific CCDs routinely achieve CTEs greater than 0.99999.

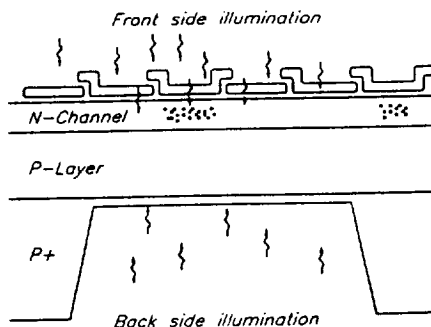
To build an imaging device, the imaging area is divided into picture elements or pixels. The image is focussed onto this region, and photons penetrate through the electrodes (made from thin silicon) and the insulator into the semiconductor. Each photon generates an electron-hole pair. The hole is repulsed, and the electron is captured due to the voltage applied to the electrode. To prevent electrons in adjacent pixels from mixing and consequently blurring the image, several electrodes are used in each pixel. At least one electrode is biased to prevent electrons from escaping into neighboring pixels. The pixels are also divided in the columnar direction by physical structures called channel stops that block the horizontal movement of electrons. The electrons are collected during an integration or exposure period that may last many minutes in the case of faint objects. Following the integration period, the CCD is now operated in the transfer manner. Referring to Fig. 2, all columns are shifted down in parallel. Upon reaching the bottom of the column, a row of charges is shifted into the horizontal register. The horizontal register is then shifted to the right to an amplifier which converts the charge signal into a voltage signal. This step often introduces the most noise into the image from the CCD. The voltage then leaves the chip and goes into subsequent amplifiers and an analog-to-digital converter. (From then on, the image is computer-compatible and can be directly loaded onto magnetic disk for later display and processing.) As each row is read out, the next row is shifted into the horizontal register. This process continues until the entire image is read out from the CCD.



**Figure 1.** Simplified illustration of CCD charge transfer for a three-phase device. Changing position of positive electrode bias causes electrons in semiconductor to shift to the right.



**Figure 2.** Schematic illustration of a portion of three-phase imager. Hatched area corresponds to one pixel. Heavy lines are channel stop structures. Arrows show path of charge transfer. Signal charge is shown confined under electrode 2.



**Figure 3.** Cross section through a scientific imager showing three-phase polysilicon electrodes, n-channel layer (typically  $0.5\mu\text{m}$  thick and doped at  $2 \times 10^{16}/\text{cm}^3$ ), epitaxial p-layer ( $10\text{-}12\mu\text{m}$  thick), and p+ substrate. Thinned structure illustrates back side illumination. Dots in n-channel represent collected signal electrons.

## **Requirements for Scientific CCDs**

The requirements for a scientific CCD are significantly different from those for home video cameras. The major differences are (1) array size, (2) frame rate, (3) dark current, (4) read noise, (5) spectral range and (6) radiation hardness. In structure, scientific CCDs are simpler than consumer CCDs since issues such as anti-blooming, color separation, and read-out rate are relaxed for the scientific CCD. In addition, cosmetic blemishes are often accepted in non-space scientific applications but are unacceptable in the consumer arena. Nevertheless, the large volume production and smaller format of the consumer CCD have pushed its unit price to a level several orders of magnitude lower than that of a scientific CCD.

### **Array Size - Frame Rate**

The array size of a scientific CCD is almost routinely 1024x1024, with sizes up to 4096x4096 demonstrated by Ford Aerospace. With a pixel pitch of 12 $\mu$ m, a 2048x2048 imager is nearly an inch square, with only a few chips produced on a 4 inch diameter wafer. The 4 million pixels in a single image frame read out at 50 kpixels/sec, and digitized to 16 bits require over a minute to read out and 8 Mbytes of computer memory to store. A 4096x4096 image requires 32 Mbytes - the size of personal computer hard drive disk.

### **Dark Current**

Normally, signal electrons are generated when incident photons generate an electron-hole pair. However, thermal excitation can also result in electron-hole pair generation. Since this occurs without optical input, it is referred to as dark current. Dark current depends both on temperature, and on defect-induced energy levels (called traps) within the semiconductor energy gap. The latter act as stepping stones for thermal excitation and can arise from starting material quality, device design, device processing, and radiation events. Dark current adds to the optically generated signal, and is noisy. In consumer video cameras, dark current is not as important due to the high TV frame rate, but in a scientific CCD using a very long exposure time (e.g. 15 min), the dark current can add up to a sizeable contribution. To avoid dark current, scientific CCDs are often cooled to -70°C or lower.

### **Radiation Hardness**

A CCD in orbit, or travelling throughout the solar system is constantly exposed to varying amounts of radiation in the form of energetic, charged particles, and high-energy photons. This radiation can damage both the insulating layer between the electrodes and semiconductor (producing unreliable operating conditions) and damage the silicon crystal. The defects caused by the radiation degrade both the charge transfer efficiency and the dark current. Unlike digital circuits that use transistors for on/off switches and are consequently more immune to such damage, CCDs suffer a gradual reduction in performance. On the other hand, transient events that cause devastating single-event upsets in digital circuits produce only one or more bright pixels in a CCD image and are not of great consequence. Hardening of CCDs to radiation damage is an important element of scientific CCD development.

### **Read Noise**

Images generated by photons are always intrinsically noisy. Photon shot noise varies

as the square root of the number of optically generated electrons. The CCD can add additional noise both in the transfer process and in the conversion from charge to voltage domains. In a scientific CCD with very high CTE, the transfer noise component can be ignored, and the charge-to-voltage conversion process noise dominates. Contributions come from the output transistor white and  $1/f$  noise. A typical read noise level for a scientific CCD is of the order of 4-6 electrons, and is larger than the photon shot noise for very faint images with less than approximately 50 electrons per pixel.

### Quantum Efficiency

For imaging faint objects, the fraction of incident photons converted into signal charge is an important parameter. Loosely termed quantum efficiency, it includes reflection loss at the front surface, loss by absorption in the electrodes, loss in the insulator, loss caused by recombination of carriers generated deep in the semiconductor, loss caused by surface recombination, and in the case of near-infrared radiation, lack of sufficient absorption in the semiconductor. In the ultraviolet where losses due to absorption in the polysilicon electrodes would be significant, two approaches can be taken. One is to coat the front side with a photoluminescent material such that the UV photon is absorbed and a longer wavelength photon is subsequently emitted into the CCD. Although an inefficient process, the quantum efficiency in the UV can be pushed from near zero to perhaps 15-20%. Alternatively, one can thin the semiconductor under the CCD from the backside. Illumination on the back surface need not pass through an electrode before becoming absorbed in the semiconductor (See Fig. 3). If the thinning is done carefully to avoid recombination losses at the back surface, the UV quantum efficiency can be increased further, at the expense of delicate material processing and handling. This is described further below. Quantum efficiency in scientific CCDs is typically better than 50% throughout the visible portion of the spectrum, with some fall-off in the blue due to absorption in the polysilicon electrodes and reflection.

### Technologies Used in JPL CCDs

JPL and its academic and industrial partners have developed technologies which specifically address the needs of scientific imagers. Three of these technologies are described here. Backside illumination technology has allowed the improvement of quantum efficiency in the blue and ultraviolet portions of the spectrum. Inversion-mode operation of the CCD has resulted in dark current reduction by several orders of magnitude. The incorporation of signal averaging circuitry in the readout amplifier has allowed read noise to be reduced to an average level of less than one electron.

### Backside Illumination

There are two components to backside illumination technology; thinning of the CCD and stabilization of the back surface. Several companies participating in thinning activities include Texas Instruments, RCA, MIT Lincoln Laboratories, Thomson-CSF, Tektronix, and EG&G Reticon. The major issues in thinning are to achieve a specularly smooth surface after etching and to stop the etch at the appropriate depth. While thinning is still far from routine, processes to produce the desired thinning have been demonstrated. Following wafer fabrication and die separation, the backside of the chip is masked except for an area directly under the imaging portion. The chip is then immersed in an acid-based etching solution. The chip is typically etched from its initial thickness of  $300\mu\text{m}$  to a thickness of  $10\text{-}12\mu\text{m}$ . At this point, the chip is thin enough to permit red light to penetrate through it. Reduction of surface stress is important to prevent the warping of the resultant thin membrane. It is also important

to minimize stress arising from packaging.

The backside surface, if not passivated, is sensitive to electrical charging effects. If it becomes positively charged, optically generated electrons are drawn to the back surface where they can recombine, reducing quantum efficiency. Ideally, the backside should be negatively charged to drive electrons toward the front surface. However, this draws positively charged holes to the backside where, if allowed to accumulate, the holes counteract the initial negative charge. Several approaches have been taken to this problem both at JPL and elsewhere. A thin layer of oxide can be used to passivate the back surface, reducing recombination sites. A very thin overlayer of metal can be used to bias the backside to maintain a constant potential. Alternatively, the shallow implantation of group III impurities can be performed to generate the required field. Optically-generated holes must still be drawn out from this structure to prevent the bleaching effect described above. The present approach for the replacement wide-field, planetary camera (WF/PC) for the Hubble Space Telescope uses the thin oxide, thin metal technology.

### Inversion-Mode Operation

Dark current in CCDs can be reduced by lowering the CCD operating temperature. A second method of reducing dark current is to operate the CCD in an inversion mode. Since defect levels at the silicon-insulator interface provide stepping-stone energy levels for electron-hole pair thermal generation, it is desirable to "plug" these paths. This can be accomplished by flooding the surface with holes, thus reducing the number of electrons available for hopping to the conduction band. Such flooding by minority carriers is called inversion. While understood for some time that such a mode of operation would reduce dark current, only recently has this technology been implemented in three-phase scientific CCDs. A special implantation step must be performed to insure that pixels remain isolated during integration and charge transfer. In one device, dark current was reduced from 30,000  $e^-/\text{sec/pixel}$  to 800  $e^-/\text{sec/pixel}$  at 30°C using the inversion mode, and the ratio remained constant down to -70°C (0.3  $e^-/\text{sec/pixel}$  and 0.008  $e^-/\text{sec/pixel}$  respectively). The inversion-mode also inhibits the increase in dark current after irradiation. For example, without the inversion mode, the dark current in the same device increased by a factor of 200 after 20 krads of Co-60 radiation, but using the inversion mode, the dark current was suppressed by a factor of 1000.

### Read Out Averaging

The read out process of converting from charge-to-voltage adds noise to the signal as described above. There are several components to the noise, and some can be decreased by increased sampling time. However, 1/f noise generated by the on-chip amplifier can increase the total noise for long sampling periods. It follows that if one could repeatedly read the signal non-destructively, one can achieve noise reduction by averaging a series of output levels corresponding to the same pixel. Such a non-destructive read out circuit was recently integrated on a Ford Aerospace imager. By non-destructively reading the signal using a floating-gate amplifier, the noise was reduced as the square-root of the number of samples taken. For example, the noise in a test pattern was reduced from 7.6  $e^-$  rms (spatial average) to 0.97  $e^-$  rms by averaging each pixel over 64 repeated samples.

The averaging process increases the read out time for an image in proportion to the number of samples taken for each pixel. For short exposure times, the increased readout time might be better utilized as a longer exposure time. For long exposures of faint objects, and for short exposures of once-only images, the on-chip averaging process can dramatically

improve image quality.

### **The CRAF/Cassini Imager**

The visible spectrum science imager under development for the CRAF/CASSINI mission represents a good example of the present state of the art in scientific CCDs. Although it is not a back-side-illuminated device, it is a large format imager with high performance.

The architecture for the imager, fabricated at Ford Aerospace, is split-frame / externally-shuttered. It is implemented with a triple-polysilicon process using a buried n-channel technology and inversion-mode implants. Its format is 1024x1024 pixels with a 12 $\mu$ m square pixel size. Each pixel can hold up to 100,000 electrons. The charge transfer efficiency of fabricated devices has been measured to be better than 0.999999 at -70°C. The read out noise at a 50,000 kpixel/sec readout rate has been measured to be approximately 5 electrons rms. The dark current for the device is specified to be under 2 nA/cm<sup>2</sup> at 22°C. When operated in the inverted mode, the dark current has been measured to be 0.025 nA/cm<sup>2</sup> at 22°C. The illumination response non-linearity for the CRAF/Cassini CCD is less than one percent, and the pixel-to-pixel response non-uniformity for white light is also better than one percent. Quantum efficiency under front side illumination is better than 45% at 600 nm. The performance goals and achieved levels are shown in Table 1.

### **Future Directions for Scientific CCD Research and Development**

There are several areas ripe for continued research and advanced development activities in scientific CCDs. For example, quantum efficiency - particularly in the blue and ultraviolet, is still low for three-phase devices. New structures which avoid the absorption of these wavelengths by the overlying electrodes can enhance imaging in these wavelengths. Back side illuminated devices continue to have stability difficulties. New techniques to stabilize the back surface are needed to achieve high quantum efficiency in the ultraviolet. Damage caused by energetic particles such as protons and neutrons remain a significant problem for CCDs. Structures and operating techniques to minimize the effect of this damage warrant further investigation.

The use of improved silicon and other materials such as GaAs and Ge offer the potential for extending the useful spectral response on the CCD. For example, x-ray imaging using high resistivity silicon has already been demonstrated. The incorporation of photo sensitive III-V layers may yield a LWIR image sensor.

The integration of additional circuitry for on-chip signal processing of the image data is another area of research. For example, investigation of parallel circuitry to perform the read out averaging so that read out rate does not suffer from the averaging process is of interest. Circuitry for performing other functions such as for sparse illumination read out might also present an avenue of investigation. Finally, the continued improvement of device processing will continue to lead to improvements in image format and array size.

The advancement of fabrication capability also presents an interesting challenge to the scientific CCD. Photodiode arrays which have traditionally had higher noise, lower fill-factors, and less uniformity than CCDs, will be able to take advantage of wafer processing improvements. These devices have higher quantum efficiency, especially at shorter wavelengths, than CCDs, and are more tolerant to radiation and processing defects. It is possible that a cross-over point will be reached for very large arrays in which photodiode arrays re-emerge as a successful competitor to CCD technology.

Table 1  
CRAF/Cassini CCD Imager

Parameter	Goal	Actual
Architecture Technology	split-frame, three phase triple-poly, n-buried channel inverted operation	
Array size	1024 x 1024	
Pixel shape	square	
Pixel pitch	12 $\mu$ m	
Well capacity electrons	>50,000 electrons	100,000 electrons
Charge transfer efficiency	> 0.99999	0.999999
Dark current non-inverted	< 2 nA/cm <sup>2</sup>	
inverted	< 0.03 nA/cm <sup>2</sup>	0.025 nA/cm <sup>2</sup>
Response non-linearity	< 1%	1 %
Uniformity	< 3 %	1 %
Read out rate	50,000 pixels/sec	
Read out amplifier sensitivity	>1.0V/e <sup>-</sup>	1.0V/e <sup>-</sup>
Read out noise	$\leq$ 10 electrons rms	5 electrons rms



**ADVANCED THERMAL CONTROL TECHNOLOGY  
for  
COMMERCIAL APPLICATIONS**

**Theodore D. Swanson, P.E.  
NASA Goddard Space Flight Center**

**ABSTRACT**

A number of the technologies previously developed for the thermal control of spacecraft have found their way into commercial applications. Specialized coatings and heat pipes are but two examples. The thermal control of current and future spacecraft is becoming increasingly more demanding, and a variety of new technologies are being developed to meet these needs. Closed two-phase loops are perceived to be the answer to many of the new requirements. This paper discusses all of these technologies and summarizes their spacecraft and current terrestrial applications.

**INTRODUCTION/BACKGROUND**

The thermal control of spacecraft presents some unique problems not generally encountered in terrestrial applications. Space is a place of great extremes: surfaces exposed to the sun can become quite hot while surfaces facing deep space can become very cold. In addition, any piece of equipment which uses electrical power must reject an equal amount of energy as waste heat. Such rejection of waste heat to the ultimate sink must be by radiation. These factors can create unacceptable temperature extremes for a spacecraft unless some measures are taken to control heat flows. In the past this control was accomplished with special coatings and a unique kind of insulation called multi-layer insulation (MLI). Somewhat later more sophisticated devices such as louvers and heat pipes were developed. These technologies have worked well for applications up to the present which involve a few kilowatts or less and only short transport distances. However, future applications such as the Space Station Freedom, Earth Observing System platforms, and Lunar Base are much more demanding. For these applications a new type of advanced thermal control technology is being developed: closed, two-phase loops which can act as a central thermal bus.

The technologies previously developed for spacecraft thermal control have eventually found their way into certain terrestrial applications. It is reasonable to suppose that the more advanced technologies now being developed will also find commercial and industrial markets. Discussed below are some selected NASA developed thermal control technologies and their spacecraft and commercial applications.

**CURRENT THERMAL CONTROL TECHNOLOGIES and APPLICATIONS**

**Coatings**

A variety of special coatings have been developed during the history of space exploration. In general, the purpose of these coatings has been to either reflect or absorb incident solar radiation, or to radiate or prevent the radiation of waste heat. These functions are characterized by the surface "absorptivity" and "emissivity", respectively, of a material. For a given surface, the ratio of these two parameters combined with the radiation impinging upon it and the heat conducted to/from it will determine the net flow of energy into and out of this surface. All surfaces taken together will determine the net thermal balance of the spacecraft. Thus, by selecting the emissivity and absorptivity of the exterior surfaces the flow of energy into and out of a spacecraft can be affected. For early spacecraft, which were small and of low power, such coatings combined with MLI were generally enough to provide adequate thermal control.

The special coatings developed for space applications have found a number of commercial applications. For example, the technology used for coating astronaut helmet visors is now used for window coatings on buildings. In addition, silicate paints, similar to those developed for spacecraft, were used during the recent

renovation of the Statue of Liberty. The aluminum oxide coatings used to protect many silver polishes from tarnishing are similar to those developed by NASA. Also, conductive coatings such as indium oxide and indium antimonide are employed to coat liquid crystals.

### Heat Pipes

Heat pipes are inherently very simple devices. They are used to conduct heat over relatively long distances with only a small temperature drop. In its conventional form, the heat pipe is a closed tube or chamber whose inner surfaces are lined with some sort of porous wick. In operation the wick is saturated with a the liquid phase of the working fluid while the remaining space is saturated with the vapor. Heat is applied to the evaporator end of the pipe. This heat is absorbed by the liquid which then vaporizes. The lost liquid in the wick is replaced via capillary forces in the wick. The resulting pressure difference between the vapor and liquid phases drives vapor from the evaporator end and to the condensing end, where it condenses back into a liquid and is absorbed by the wick. Capillary forces in the wick pump the liquid back to the evaporator area. Hence, a closed heat transport loop is created. This process is depicted in Figure 1. since the process makes use of the fluid's latent heat of vaporization a significant amount of heat can be transported relative to fluid flow rate and unit size. The process can continue indefinitely as long as the fluid flow passage is not blocked and a sufficient capillary head can be maintained. There are no moving parts to wear out.

Heat pipe technology was originally developed in the early 1960's. They have since been extensively used for spacecraft thermal management. A wide variety of heat pipe designs have been developed over the years to meet different thermal control needs. However, almost all of these designs were for moderate temperature applications (e.g., near room temperature) and employed ammonia as the working fluid. Due to basic thermophysics a given heat pipe design will be applicable only for a limited temperature range. Ammonia, which has both a very high latent heat of vaporization and a high surface tension (which is needed for good wicking) is an outstanding fluid for moderate temperature applications from about minus 40 degrees Celsius to about plus 50 degrees Celsius. Although a few experimental heat pipes for cryogenic applications have been fabricated, these are not yet generally available for commercial or spacecraft applications. High temperature heat pipes have, however, been fabricated and are now commercially available. These pipes employ a liquid metal as the working fluid.

Heat pipes normally operate as high efficiency heat transfer devices with a fixed conductance. However, by modifying their design slightly it is possible have them operate with a conductance that is variable in response to some external driver function. It is also possible to design a heat pipe so that it operates as a diode and conducts heat only in one direction.

Although deceptively simple in concept, development of a properly operating heat pipe requires careful consideration to numerous design details and the cleaning and charging procedures. Special procedures must also be employed in testing. References 1 and 2 provide a good summary of the basic issues involved in heat pipe design.

Heat pipes have found their way into a wide variety of commercial applications. Virtually all commercial communication satellites depend upon heat pipes for thermal control. There are also a large number of terrestrial applications. For example, elements of the technology were used for thermal control of the perma-frost for the Alaskan pipeline (over 125,000 pipes were used). In addition, heat pipes have been used in many industrial waste heat recovery systems. Many high temperature furnaces use heat pipes to provide extremely isothermal conditions (less than 1 degree Celsius) . Heat pipes are also used to remove heat from industrial and medical (e.g., dental) drill bits. A major application is for cooling high power electronic equipment, especially in military aircraft.

The primary limitation to the use of heat pipes in terrestrial applications is generally not power or transport length. Rather it is typically the maximum elevation that can be tolerated between the evaporator and the condenser ends. Generally, the evaporator can be no more than a few inches above the condenser or the wick cannot pump and the heat pipe will fail. However, if the condenser can be on top then the pipe can

be operated in a "reflux" mode and large height differentials are possible.

### Two-Phase Pumped Loops

Future planned space structures, such as the Earth observing System platforms and Space Station Freedom, will pose a much more demanding thermal control problem than current spacecraft. They will have power levels in the tens of kilowatts, transport distances in the hundreds of feet, numerous heat load sources dispersed throughout the structure, and will require tight temperature control ( $\pm 2$  or 3 degrees Celsius). In addition, the numerous heat load sources will all require this cooling under a variable load schedule. These requirements virtually prohibit the use of conventional passive or single phase thermal control technology.

Pumped, two-phase thermal control technology has been developed to meet the emerging requirements discussed above. This technology is essentially a major evolutionary step beyond heat pipes and offers about a two order magnitude improvement. In a two-phase loop, a subcooled liquid refrigerant (ammonia is the preferred fluid) is first introduced into an evaporator, which typically has some sort of wick structure to promote fluid distribution. Heat from the equipment to be cooled is applied to the evaporator where it is absorbed by vaporization of the liquid. It is important to note that the equipment to be cooled sees an isothermal sink. The refrigerant then leaves the evaporator as either a saturated vapor or two-phase fluid. It then travels through a tube to a condenser where it gives off heat and returns to a liquid state. This condenser may be either a radiator or a heat exchanger connected to some intermediate sink. The process is depicted in Figure 2. References 3 and 4 provide a brief technical overview of the technology.

Three basic types of closed, two-phase loops have been developed. These are characterized primarily by the type of motive force used to circulate the fluid through the loop. In a Capillary Pumped Loop (CPL) the liquid is circulated by only the wicking forces generated within the evaporator's wick. Figure 3 depicts such a process. This is similar to a heat pipe, except that in a CPL the liquid and vapor phases are separated and flow through different lines. The effluent from such capillary evaporators is typically a saturated vapor. Outstanding characteristics of this system type include the fact that the capillary based evaporators are self regulating (i.e., the more power applied, the more they pump, up to their limit). In addition there are no moving parts to wear out. A potential limitation is the relatively small pressure head generated (half of a psi is typical). CPL's have demonstrated a two order of magnitude improvement in heat transport capability (power times distance) over heat pipes.

The second major type of two-phase loop involves the use of a small mechanical pump to supply refrigerant to the evaporators. In this type of system, depending on the system design the evaporator may discharge either a pure vapor or a vapor/liquid mixture. The evaporators, however, do not provide any pumping action of their own. In either case the effluent goes to a condenser (heat exchanger or radiator) and is condensed back into a liquid. It is then pumped back to the evaporator and the cycle continues.

In the third major system type, a mechanical pump is used in concert with capillary evaporators. This combines many of the advantages of the capillary and mechanically pumped systems; the capillary pumps are self regulating and the mechanical pump provides additional pressure head to drive the fluid through the loop. This system concept is depicted in Figure 4. Variations on these basic system concepts are also possible.

Two-phase thermal control technology has been baselined for both the Earth Observing System platforms and Space Station Freedom. These large space facilities will require the high power capacity, long transport capability, low relative weight, and isothermality offered by two-phase technology. It is anticipated that any future large space facility with similar requirements would also need to make use of this technology.

To date there are no known commercial applications of the two-phase thermal control technology discussed above. Ammonia refrigeration systems have been operated in industrial and commercial applications for many years, but these have involved the generation of a temperature differential for refrigeration purposes. The technology discussed above is concerned with the efficient transfer of heat over relatively long distances, not refrigeration. However, just as heat pipes have proved useful for commercial

applications it would appear that two-phase loops could solve some terrestrial heat transfer problems. The use of a mechanical pump might eliminate the height restriction problems existent with heat pipes.

### **FUTURE THERMAL TECHNOLOGY**

The technology discussed above will be able to provide thermal control for the NASA missions scheduled to date. However, several additional technologies will be required to enable or significantly enhance future proposed space activities. For example, in order to provide thermal control for a Lunar Base a new type of heat rejection or thermal storage technology will be required. Heat pumps have been proposed to solve this problem. Although heat pumps have been used for years in terrestrial applications some modifications will be required for space applications. As these modifications will likely address reliability and efficiency some technology transfer back to the commercial sector seems reasonable. Other new thermal control technologies needing development for space applications include low temperature heat pipes and thermal storage devices. If the past is any guide then these technologies will also offer terrestrial applications.

### **REFERENCES**

1. Brennan, P.J. and Kroliczek, E.J. ; Heat Pipe Design Handbook. B & K Engineering, Inc., prepared for NASA Goddard Space Flight Center, 1979.
2. Chi, S.W. ; Heat Pipe Theory and Practice: A Sourcebook. McGraw-Hill Book Company, 1976.
3. Kroliczek, E.J., Ku, J., and Ollendorf, S.; Design, Development and Test of a Capillary Pump Loop Heat Pipe. AIAA 19th Thermophysics Conference, 1984.
4. McCabe, M.E. and Ku, J.; Design and Testing of a High Power Spacecraft Thermal Management System. NASA Technical Memorandum 4051, 1988.

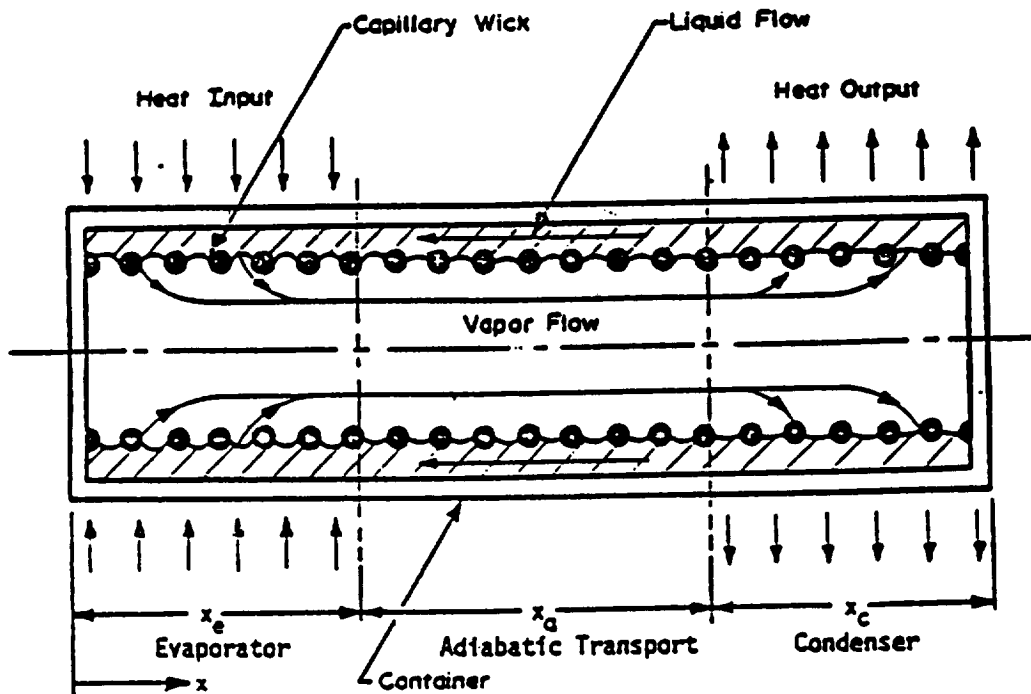


FIGURE 1. Components and Principle of Operation of a Heat Pipe

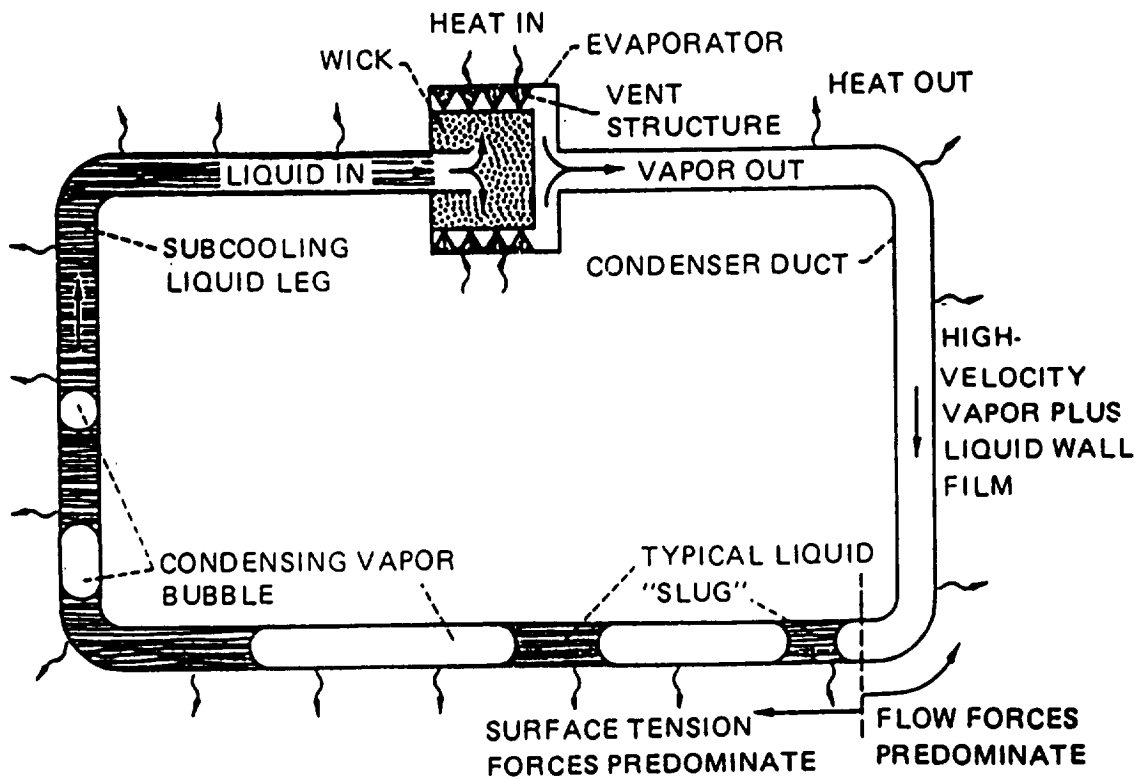
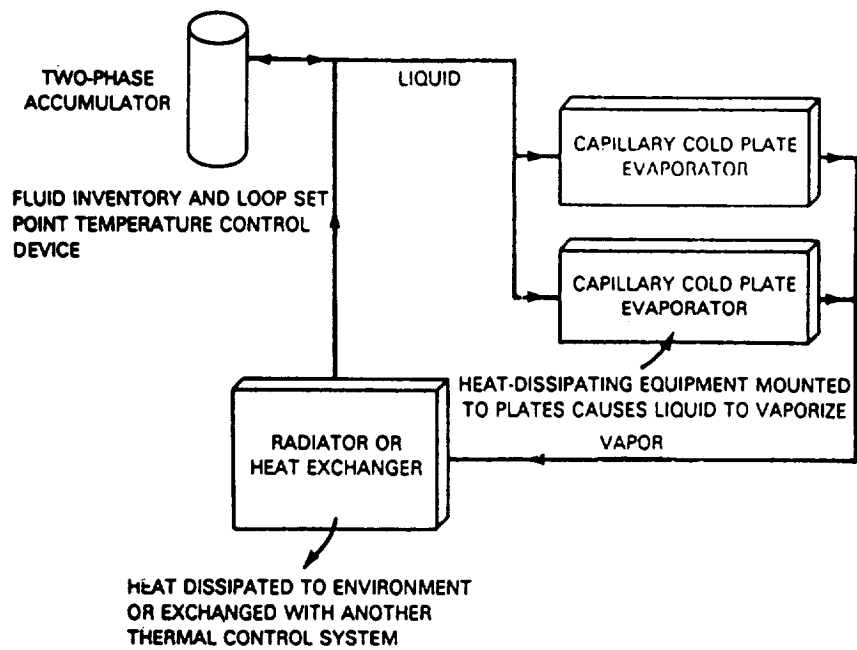
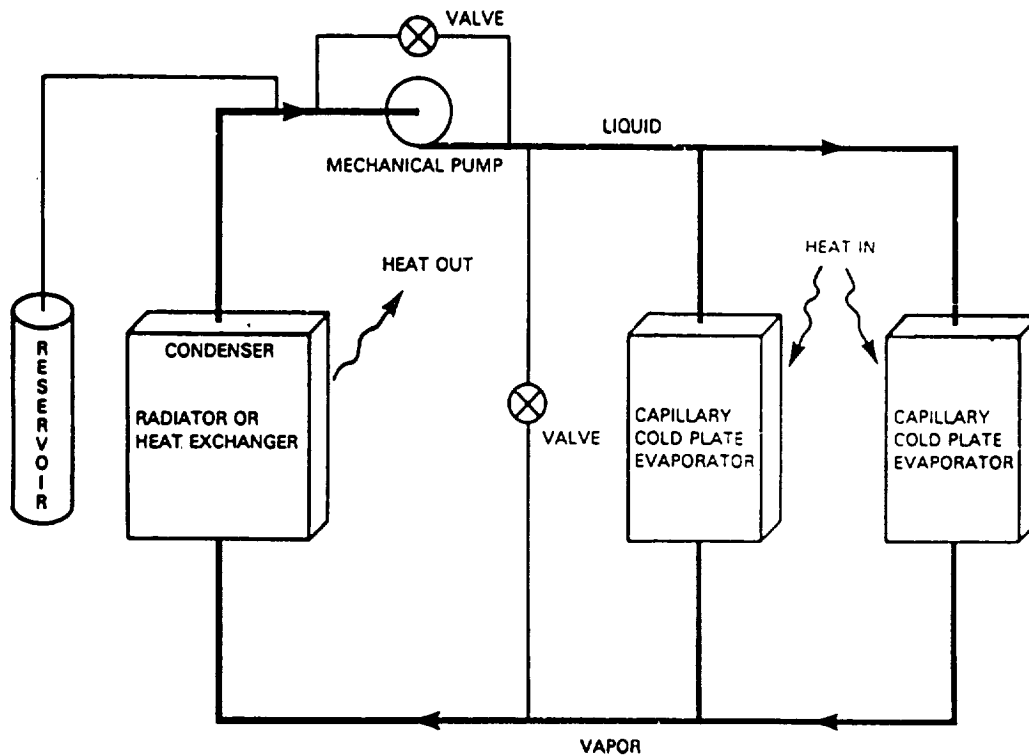


FIGURE 2. Principle of Operation of a Closed, Two-Phase Loop



**FIGURE 3. Capillary Pumped Loop**



**FIGURE 4. Capillary/Mechanically Pumped Loop (Hybrid)**

## **SESSION G - ROBOTICS**

**Tuesday November 27, 1990**

- **ROBOSIM: A Simulator For Robotic Systems**
- **Control Systems Software, Simulation, And Robotic Applications**
- **Telerobotic Electronic Materials Processing Equipment**
- **Advanced Mechanism For Robotics**
- **The Flight Telerobotic Servicer And Technology Transfer**
- **FARMS: The Flexible Agricultural Robotics Manipulator System**
- **Diverse Applications Of Advanced Man-Telerobot Interfaces**





**ROBOSIM, A SIMULATOR FOR ROBOTIC SYSTEMS**

**Elaine M. Hinman and Ken Fernandez, Ph.D.**  
**Marshall Space Flight Center**

**George E. Cook, Ph.D.**  
**Vanderbilt University**  
**Nashville, Tennessee**

ROBOSIM, a simulator for robotic systems, was developed by NASA to aid in the rapid prototyping of automation. ROBOSIM has allowed the development of improved robotic systems concepts for both earth-based and proposed on-orbit applications while significantly reducing development costs. In a cooperative effort with an area university, ROBOSIM has been further developed for use in the classroom as a safe and cost-effective way of allowing students to study robotic systems. Students have used ROBOSIM to study existing robotic systems and systems which they have designed in the classroom.

Since an advanced simulator/trainer of this type is beneficial to not only NASA projects and programs but industry and academia as well, NASA in the process of developing this technology for wider public use. In this paper we give an update on the simulator's new application areas, the improvements made to the simulator's design and current efforts to ensure the timely transfer of this technology.

**INTRODUCTION**

ROBOSIM is a computer graphic robotic simulator developed by NASA-MSFC to aid in rapid prototyping of automation concepts. Cooperative development of a commercial and classroom version of ROBOSIM has been carried out by MSFC and Vanderbilt. In addition, ROBOSIM has been used in graduate courses at Vanderbilt as a trainer in robotic design.

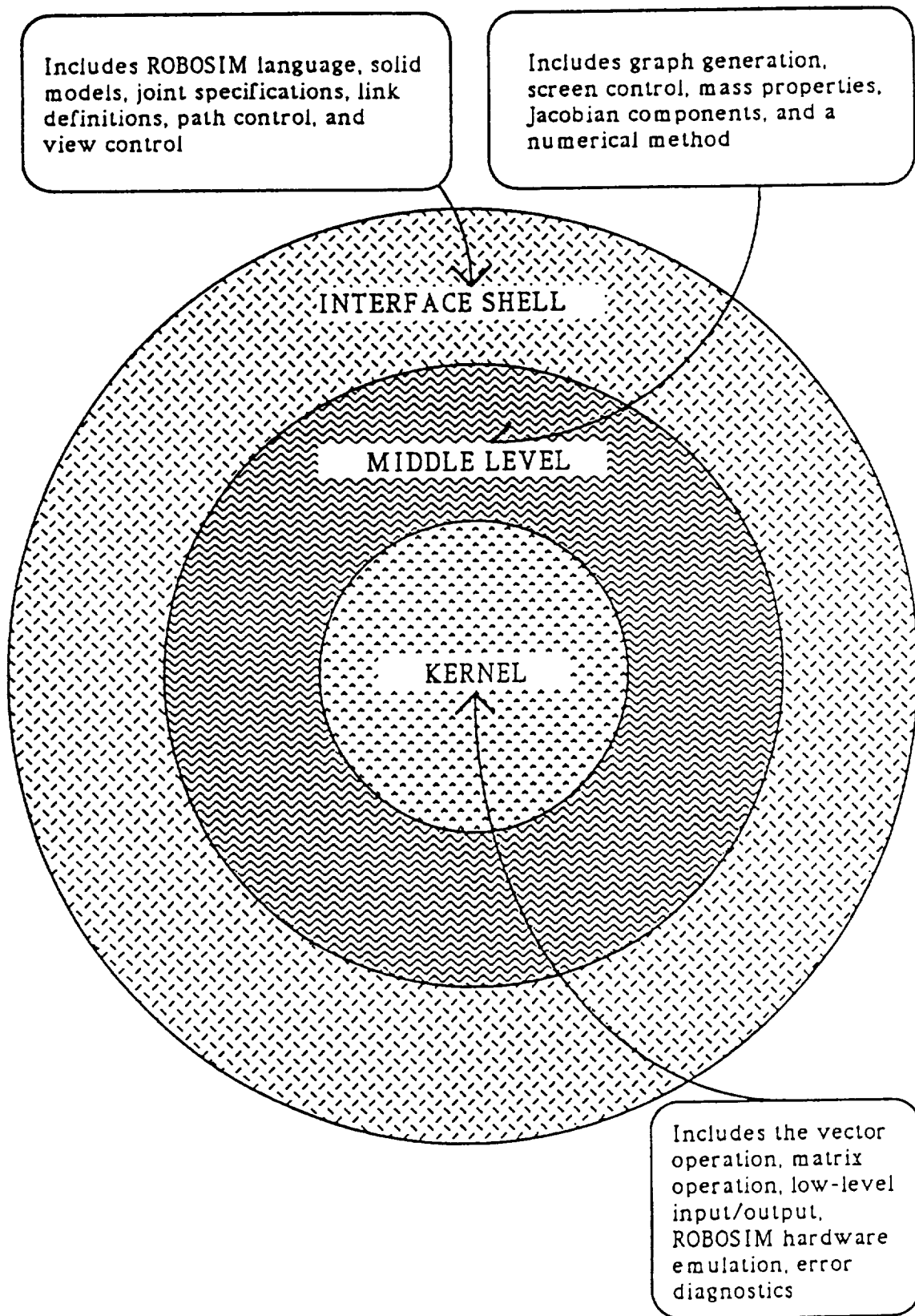
Several applications of ROBOSIM will be described in this paper. While these are not the only current uses, they do show the usefulness of ROBOSIM in diverse fields -- from spaceflight to manufacturing to surgical aid development.

In the course of making ROBOSIM accessible to the larger robotics simulation and development community, several ports have been made to other workstations. A couple of these are described in more detail below.

**DEVELOPMENT**

ROBOSIM was originally developed in a VAX environment and provided graphical output to TEKTRONIX 4014 terminals and Evans & Sutherland graphics terminals [3]. Ports to the GTI Poly 2000, HP350SRX workstation, InterPro 360, and Silicon Graphics IRIS machine also exist [1, 2]. The TEKTRONIX terminal provides a low-cost and quick way to see the robot and workcell during the creation process, while the other systems allow real-time or near real-time viewing of robot motions.

To use the original ROBOSIM, the user created shapes and interconnecting joints using the ROBOSIM programming language. once a robot was modelled, motions could be generated either by user supplied routines or by using the default control algorithms [4]. The ROBOSIM program structure is shown in Figure 1.



**FIGURE 1. Tri-level structure of ROBOSIM**

## PORTS

The advent of high speed graphics workstations provided the impetus to port ROBOSIM to such a workstation [6]. However, to remain compatible with the VAX version of ROBOSIM, the VAX version has been kept as the kernel, with various features added to the front and back ends.

The TEKTRONIX 4014 terminal and Evans & Sutherland PS330 graphics system only allowed wireframe representations of the robot and workcell. The second implementation of ROBOSIM was on an HP350SRX graphics workstation. This allowed solid graphics, faster screen access, and facilitated the development of the R2 graphical interface discussed in the next section. The Starbase graphics library and X-Windows capabilities available on this workstation were used in the implementation of pull down menus and shaded graphics in the new ROBOSIM user interface.

A parallel development has been to integrate ROBOSIM onto the Intergraph InterPro 360 graphics workstation [2]. The Intergraph windowing environment, Environ V, was used in interfacing ROBOSIM with the InterPro workstation. Again, the foundation of ROBOSIM remained the same, however, the ROBOSIM-created graphical libraries were generated using Intergraph's graphic capabilities, and displayed in an Environ V window.

## GRAPHICAL EDITOR

Once ROBOSIM was ported to the HP workstation, the 3D graphical editor, R2, was developed to provide an easier and more flexible interface to ROBOSIM [3]. The ROBOSIM kernel handles the rudimentary simulation tasks including the generation of primitive solids and rotation and translation operations performed on those objects. When modelling using the ROBOSIM programming language, the user was required to keep track of each object's dimensions, position and orientation with respect to other objects it may be connected to. In R2, menus and a mouse are used to generate the link primitives and to define the interconnecting joints. Since R2 is a graphical editor, the user may quickly and easily see and modify the robot workcell as it is being constructed. However, complete compatibility with ROBOSIM is maintained by having R2 generate ROBOSIM code. This way the capabilities of the original ROBOSIM are maintained while the enhancements of R2 are provided. Now, the user can design robots in the ROBOSIM code, by using R2, or through the use of custom programs.

## SIMULATION LIBRARY

The following description of the ROBOSIM simulation library is from the paper by Springfield, Cook, Anderson, and Fernandez [3]. The library was first implemented on the HP-9000 workstation equipped with an SRX graphics accelerator.

The simulation library and environment provides methods to access the data structures created by ROBOSIM. The robots and other objects are specified and loaded into memory. These structures remain resident in memory while the simulation is running. The library provides an interface to these structures so that the user does not have to understand what is happening at that level. The library provides higher level facilities much like an actual robot programming language.

The simulation package allows one to use the robots that have been designed. The package consists of a library of C functions that operate on the files created by ROBOSIM. Although this package is far from complete, it allows simple simulations to be run. It also provides a framework in which to test the major components for the simulator: collision detection and dynamics. Having the simulator as a library of C routines allows more flexible methods for running simulations. Very specific and efficient simulations can be written in C which call the simulation functions directly. However, even at this level, much of the internal data structure is hidden from the user. This level of programming roughly corresponds to programming a robot in its programming language. For instance, one can tell a robot to move along a straight line or move a particular joint. Using these same routines, a very flexible, user-friendly interface can be built up, allowing an

interactive way to do simulations that are not too complicated, or that do not require great speed.

ROBOSIM provides most of the information required by the simulator through the files it creates. However, some information is not directly provided, but it can be determined from what is there. This involves the information required by the collision detection algorithm. ROBOSIM provides the Denavit-Hartenberg parameters, the A-matrix, the pseudo-inertia matrix, and a list of points which describe the physical structure of the robot. The internal data structure also includes areas that are not currently used, but will be at a later time. These include minimum and maximum joint angles, velocities, and accelerations. The structure also includes information related to the graphics display. The simulation package acts as intermediary between the user and the internal representation.

The simulation program that the user writes can turn on collision detection, request solutions to inverse kinematics problems, and display results graphically. The user can use the general numerical Jacobian method for inverse kinematics or provide an exact solution for the robot. The user simply passes the address of the function to the simulator, and the simulator will then use that function when solving inverse kinematics for that robot. A proposed extension to the simulator will allow the recognition of the possible robot configurations for which exact solutions exist. The exact solutions to these configurations would then be used instead of a numerical method, freeing the user from having to solve and code it himself. Several uses of the simulations system can be found in later sections.

Collision detection is very important in simulation of robots. one usually wants to know if the robot has collided with its environment or with itself. The collision detection algorithm implemented here is very similar to the POCODA (Polygon Collision Detection Algorithm) algorithm given in Scott E. Walter's dissertation from Cornell [7]. The collision detection algorithm has only two weak points. It does not handle concave polygons, and it will not signal a collision if one object is completely inside of another. The stipulation concerning concave polygons is not serious. ROBOSIM does not generate concave polygons unless they are the result of a custom object. Although R2 does not check for concave polygons, this feature could be implemented. In fact, algorithms exist to split concave polygons into convex polygons. Either of these features could be implemented fairly simply. The problem of not detecting a collision if one object is completely inside of another derives from the fact that the algorithm used is a polygonal collision detection algorithm and not a solid object one. However, assuming two objects start off outside of each other and movements are sufficiently small, then this should not prove to be a problem. This condition also prevents the ability of one object to pass through another (i.e. a movement is large enough that two objects do not overlap at any point). This algorithm does not detect collisions in the volume swept by an object moving between positions with another object, but rather only overlap of the objects at the starting and ending positions. But, if the distance between the positions is smaller than the smallest object, then there should be no problems.

## APPLICATIONS

ROBOSIM has been used in several different applications. As a rapid prototyping robotic simulation system it is especially applicable when designing custom robotics for specific tasks. However, standard industry robots can also be modelled. Control algorithm development, positioning and task viewing studies, and workcell configuration are tasks for which ROBOSIM has been used.

## CLASSROOM

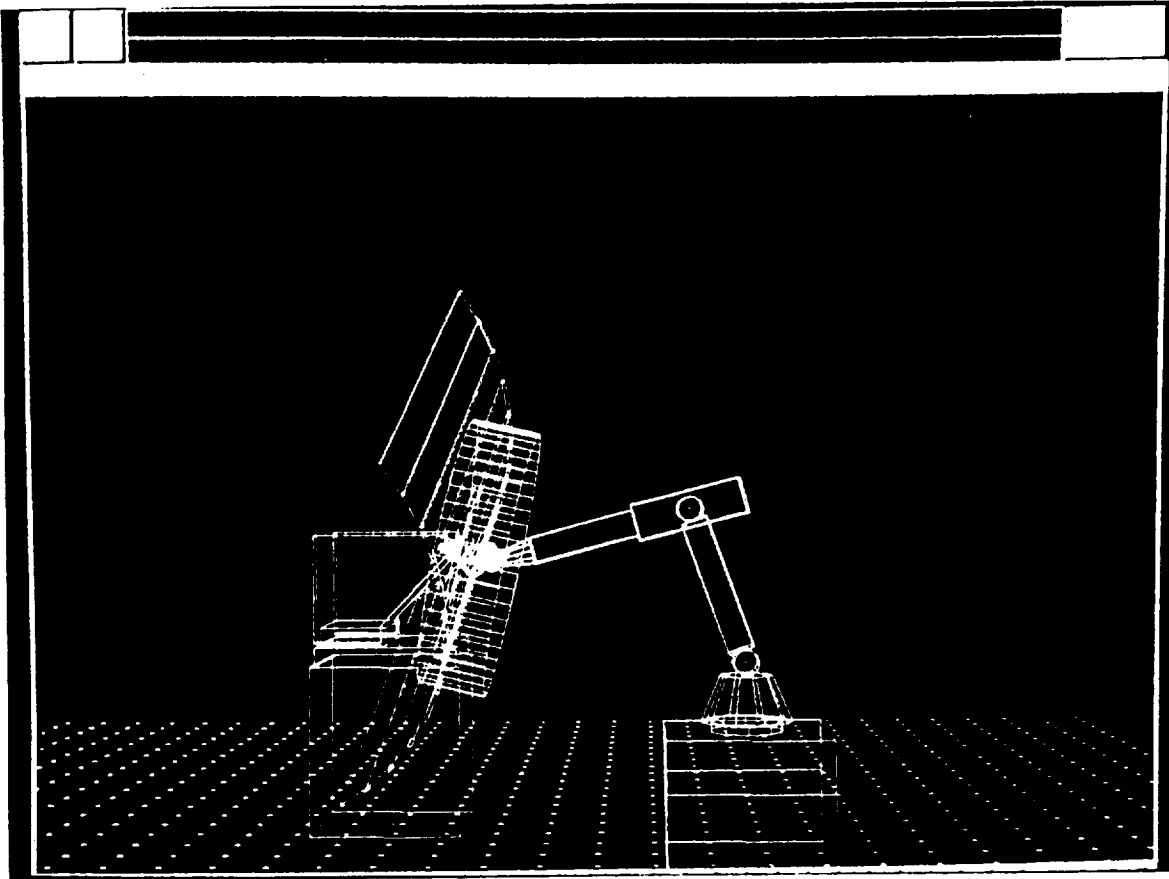
ROBOSIM is used in an introductory graduate-level robotics class at Vanderbilt University. It is useful for aiding the students in visualizing transformations and perspective operations applied to items they have designed themselves. They can build objects and link them into kinematic chains. Once a robot is designed, the inverse kinematics can be solved and various control algorithms tested in graphics. Using graphic simulation, students have access to robotics development tools without the expense of providing robotics hardware for a classroom/laboratory environment.

## **SURGICAL POSITIONER DEVELOPMENT**

ROBOSIM has been used by Dr. Bob Galloway and Allison Balogh of Vanderbilt to aid in development of a surgical positioner for use in brain surgery [3]. The positioner would be used to determine points of entry into the brain. Currently this is done by precomputing the location of the points with respect to an external reference. Greater flexibility would be provided by having a way to immediately see these positions. The current research is to see if a robot of sufficient accuracy can be built. ROBOSIM has been used to design the positioning arm and study its work envelope. The robot needs to be able to reach all positions on the head, while avoiding collisions with it. Currently, different positioner arm configurations are being studied.

## **WELDING**

ROBOSIM has been used in a couple of welding simulation applications. one was in the development of a control algorithm for robotic welding of a Space Station airlock [2]. This application was developed on the InterPro workstation. A wireframe representation from this application is shown in Figure 2.



**FIGURE 2. Airlock and Robot during weld.**

Another welding application was in the testing and simulation of a downhand welding algorithm [5,8]. In this case a six degree-of-freedom robot performs the welding while a two degree-of-freedom positioner holds the workpiece. One implementation of this algorithm used a 2 degree-of-freedom table as the positioner, while another implementation used a PUMA 560 with 4 locked degrees of freedom as the positioner. In both cases, the object is to keep the weld feed in the downhand position to avoid pooling of the field. Based on the status of objects in the workcell, path information is generated and simulated by ROBOSIM in order to perform the weld. Joint constraints and collision detection are included in the path generation. Once a path is found, it may be sent to the actual robots and executed.

## CONCLUSIONS

The previous examples point out a few of the areas in which the use of ROBOSIM is beneficial. Although ROBOSIM was originally developed at NASA for use in space robotics, it is useful to the larger robotics community. The commercial version currently under development will include the enhancements made in the second and later implementations of ROBOSIM. In this version, X-Windows may be used by multiple users to develop their applications. Although the processing speed of a workstation will be required for real-time and near-real-time simulation runs, workcell development can be carried out at a terminal through the common X-Windows interface.

In closing, ROBOSIM is a useful tool for robotics simulation in academia, government and industry.

## ACKNOWLEDGEMENTS

The authors would like to express their appreciation to the Technology Utilization Office at MSFC, and in particular to Mr. Ismail Akbay for his support of this work.

## REFERENCES

- [1] Fernandez, K., R., and E. M. Hinman, "The Use of Computer Graphic Simulation in the Development of On-Orbit Tele-Robotic Systems," Proceedings of SPIE--The International Society for Optical Engineering, Advances in Intelligent Robotics Systems, 1986.
- [2] Wilson, S., L., Interfacing of a Robot Simulation Program with Graphic Utilities of an Intergraph Interpro 360 System, Master's Thesis, Vanderbilt University, 1990.
- [3] Springfield, J., G. E. Cook, K. Andersen, and K. R. Fernandez, "ROBOSIM: A Simulation Package for Robots," University Programs in Computer-Aided Engineering, Design, and Manufacturing, ASCE Seventh Annual Conference, July 23-26, 1989.
- [4] Fernandez, K., R., "ROBOSIM User's Guide," Internal Publication of the Electrical Engineering Department, Vanderbilt University, Nashville, TN, July 1986.
- [5] Fernandez, K., R., Robotic Simulation and a Method for Jacobian Control of a Redundant Mechanism with Imbedded Constraints, Ph.D. Dissertation, Vanderbilt university, 1988.
- [6] Springfield, J., ROBOSIM Workstation Extensions, Master's Thesis, Vanderbilt University, 1988.
- [7] Walter, Scott Edward, Polygonal Collision Algorithm, Ph.D. Dissertation, Cornell University, 1985.
- [8] Fernandez, K., R., and G. E. Cook, "Computer Graphic Simulation of an Algorithm for Controlling Downhand Position in Robotic Welding," SME paper MS86-209, Presented at the SME Conference on Robotic Solutions in Aerospace Manufacturing, March 3-5, 1986.

**CONTROL SYSTEM SOFTWARE, SIMULATION AND ROBOTIC APPLICATIONS**

by **HAROLD P. FRISCH**  
**Head, Robotics Applied Research**  
**Goddard Space Flight Center**

**ABSTRACT**

The Goddard Space Flight Center (GSFC) has developed several computer programs that automatically derive and numerically solve the equations of motion for any multibody system. That is, any system that can be modeled as a collection of hinge connected rigid and flexible bodies acted upon by internal and external loads. During the past decade, these programs have proven to be the only logical method for reducing the cost of in-depth analysis for spacecraft, robots, and other complex mechanical systems. This paper provides an overview of associated modeling capability and its many applications. Several ongoing efforts to enhance existing capabilities and to apply evolving technology in and out of the aerospace industry are discussed. The man/machine interaction dynamics and performance (MMIDAP) project is a prime example. It is focused toward supporting designers of mechanical systems that are controlled by an operator's intelligent physical exertions. Medical industry involvement in this project is in a very real sense proving that technology transfer is a two way street. Many of MMIDAP's human performance and biomechanical analysis capabilities are a direct result of collaborations with the medical industry.

**INTRODUCTION**

The Technology Utilization Office, the Flight Telerobotic Servicer (FTS) project and the Office of Aeronautics, Exploration and Technology are all providing support for projects that utilize the capabilities of multibody modeling tools. This diverse support group has created a product development environment that is supportive of near-term NASA applications, far term NASA needs, and technology transfer initiatives into non-traditional application areas.

During the early 1960s, NASA satellite designers were conceiving and building spacecraft with innovative controllers faster than project support engineers could derive, code, and debug system stability and performance analysis programs. In an attempt to keep pace with need and to reduce associated cost, several efforts were initiated within the aerospace community to create a general purpose design and analysis capability. While several groups developed in-house propriety codes with impressive capability, NASA/GSFC created and brought the programs NBOD & DISCOS into the public domain. A recent review of these programs and their current capabilities is to be found in [1]. Unknown to the aerospace community, at the time, researchers at the University of Iowa were also actively engaged in developing similar capabilities for the mechanical engineering community. While GSFC supported the aerospace community, Iowa supported the machine and vehicle design community. The software of both groups is used internationally. Both development groups are now collaborating to enhance their respective capabilities. This enhancement effort is lead by a research team associated with the National Science Foundation (NSF), U.S. Army Tank Automotive Command (TACOM), NASA/GSFC cosponsored Industry/University Cooperative Research Center (I/UCRC) at the University of Iowa and its supporting membership of about 30 government and industrial laboratories. The research objectives of the I/UCRC are defined in [2].

One major trust of the collaborating researchers is to provide a reliable modeling and simulation capability for supporting the concurrent engineering systems that are presently being implemented throughout this nation's manufacturing industry. This activity is motivated by the general agreement that about 70% of a product's life-cycle attributes are defined by decisions made during the concept exploration phase of design. As a result, early in depth analysis is critical to impacting design decisions. Design changes made after this phase have a high ripple effect on project cost.

An analysis capability that can quantify operator performance vs. design alternatives during the early design phase is needed for both ground and space based applications. The GSFC's MMIDAP project supports this goal. It is directed toward machines that are controlled by a human operator's intelligent physical exertions. These tools will allow designers to quantitatively introduce an operator's physical and cognitive limitations into design tradeoff decisions (for more detail see [3] and [4].) GSFC's supported MMIDAP capability will be developed in a generic manner so that it can be applied to a broad range of aerospace, machine design, ergonomic, physical therapy and rehabilitation engineering problems.

### **NASA NEEDS, STATE OF THE ART**

The final report of the 1985 Integrated Ergonomic Modeling Workshop [5] contains a detailed review of pre-1988 software capability along with a list of recommendations for future research. It specifically remarks that "there is a paucity of dynamic interface models" and that "an integrated ergonomic model is needed, feasible, and useful." The report's review of existing capability demonstrates that ergonomic modeling software has been primarily developed to support aerospace cockpit design, design for product maintainability, and whole body dynamics associated with automotive vehicle crash and pilot ejection. Some work exists under the general heading of optimization of sports motion. However, there is virtually nothing to support designers who must evaluate man/machine interaction dynamics and performance with or without survival gear, in hostile environments, on earth, or in the reduced gravity environment of space.

A MMIDAP analysis capability is of critical importance to NASA. Space Station Freedom (SSF) is to be a long life mission. External maintenance of the SSF will be required for several decades. What cannot be accomplished via robotics will have to be done via extravehicular activity (EVA). As identified in the S-SF Fisher - Price report [6] a major goal is to reduce EVA activity by exploiting the fleet of U.S., Canadian, and Japanese robots for carrying out the maintenance and replacement of on-orbit equipment. For this strategy to be effective the orbital replacement units ORU's must be designed for commonality in order to maximize the use of robotics. The cost prohibitive need to laboratory test replacement scenarios for each of the approximately 450 different ORU types must be minimized. The development and use of a MMIDAP analysis capability is, to this author, the only viable option.

### **ANTHROPOMETRIC, BIOMECHANICAL, AND HUMAN PERFORMANCE DATABASES**

The MMIDAP project research group feels that it is time to take stock of present analysis methods and their associated data determination needs. A dedicated emphasis on developing the infrastructure for the systematic, engineering approach to solving human system problems and recognizing current limitations is needed.

The National Library of Medicine (NLM) is currently undertaking a first project at building a digital image library of volumetric data representing a complete normal adult human male and female [7]. This "Visible Human Project" will include digital images derived from photographic images from: cryosectioning, computerized tomography, and magnetic resonance imaging. NASA/GSFC is complementing this effort with a "whole body digital mapping project". The GSFC project seeks to develop a hierarchical tree of biomechanical and human performance analysis capability vs data availability. The output of this project will be used by the MMIDAP project to both define objectives and recognize analysis feasibility limits. This is a cooperative project involving: biomechanics groups from the University of Iowa and Case Western Reserve University, human performance specialists from the University of Texas at Arlington, and anatomists from the University of Colorado at Denver. The objective is for the analysts to define data needs while anatomists are to define if it is feasible with modern technology to provide the requested data as a by-product of the "Visible Human Project." Details will be defined in the project's final report [8].

One major problem with existing biomechanical data is that it comes from so many different sources with almost as many different measurement reference frames. The NLM's Visible Human Project is presenting the biomechanics community with a unique opportunity to fill gaps and to obtain a consistent reference source of fundamental biomechanical data. GSFC's whole body digital mapping project seeks to precisely define what



should be measured and how it is to be databased.

Biomechanical data alone is not sufficient for man/machine interaction dynamics and performance analysis. The problem also needs human function and human performance information. Reference [9] provides a review of ongoing work in the quantitative measurement, assessment, and databasing of human performance at the Human Performance Institute (HPI) at the University of Texas at Arlington. This work was originally focused toward the field of Physical Therapy and Rehabilitation Engineering. It is now recognized that HPI's databased information and measurement systems are identical to what is needed to support the MMIDAP project.

### **MAN-IN-THE-LOOP SIMULATORS FOR COMPLEX MECHANICAL SYSTEMS**

Major advances in formulating the mathematical equations needed to simulate complex mechanical equipment along with the availability of low cost parallel processor computers have provided a unique opportunity to create low cost real-time simulators for complex mechanical equipment with man in the control loop. Simulators accept real-time man in the loop commands, graphically create a simulated visual environment, and drive other laboratory devices to create a simulated vibrational and audio environment. Stress and load information for machine components can be obtained directly from the simulator. Qualitative control system feel information can be obtained from operator comments. Human performance data can be obtained by monitoring operator response in the simulated environment. The ability to simulate in real-time all gyrodynamic loads and machine force feedback control loads that operators must respond to sets this new effort apart from that available via aircraft flight trainer technology.

A first step toward developing a simulator capability for complex human operated mechanical systems was taken at the University of Iowa with the development of a simulator for a J.I. Case backhoe. The ongoing second step is to create the Iowa Driving Simulator in 1991 [10] and the final step will be to develop the Department of Transportation's National Advanced Driving Simulator [11] in the mid 1990's.

### **INTEGRATED MUSCULO-SKELETAL MACHINE DYNAMICS EQUATIONS OF MOTION**

The exact same methods and computer programs that are used to create real-time man in the loop simulators for mechanical systems can be used by biomechanical groups to simulate human body response. An overview of existing pre-1990 multibody modelling capability is provided in [12]. In the late 1980's several international groups independently discovered that equations of motion could be rederived in such a manner that computational speed could be greatly enhanced. New implementations with improved speed and modeling capability are now in the beta-site testing stage. The GSFC/Cambridge Research program Order N DISCOS is defined in [13] and Iowa's I/UCRC Order  $N^2$  program NGDC is defined in [14]. For biomechanical applications several modeling limitations have been recognized by Frisch, Turner, and Chun. Plans are now underway to enhance Order N DISCOS program accordingly.

Multibody simulation models have been successfully used to model certain classes of musculo-skeletal systems. As stated above, modeling weaknesses do exist and these must be recognized before one attempts to use multibody tools for general biomechanical application. The dynamics analysis of mechanical systems is dominated by the need to solve the forward dynamics problem. That is, given a prescribed set of internal and external loads, predict system response. Attempts to perform forward dynamics analysis with neuro-musculo-skeletal systems is usually stopped by one's inability to mathematically characterize the human's cognitive processes that generate the neural activation signals that stimulate the body's musculo actuator system.

The MMIDAP project recognizes this fundamental limitation, instead it concentrates on the inverse dynamics problem. Graphical animation and laboratory testing techniques exist for obtaining an estimate of human dynamic response for a broad range of activities. If sufficient information can be obtained

(displacement, rate, acceleration, and external loads) then inverse dynamics methods can be used to predict what the resultant musculo actuator loads had to be to produce the defined input response. These results can then be compared with known human performance information to determine if predicted musculo response required by a machine operator is within the limits of capability for the machine designer's target operator population group.

### **MUSCLE MODELING AND LOAD SHARING**

Detailed neuro-musculo-skeletal modeling of the human system or any of its subsystems is an extremely complex problem that is beyond today's state-of-the-art capability. The First World Biomechanics Congress in August 1990 had over 80 oral presentations on the subjects of multiple muscle systems, biomechanics and movement organization. Formal reports on 46 of these presentations have been collected in [15]. From these reports and others presented at the Congress it is clear that muscle dynamics and neuro-musculo-skeletal organization and movement modeling is a subject that will occupy researchers for many more years.

Complexities associated with modeling muscle contraction dynamics are matched by the problem of muscle load sharing. The presence of redundant muscle actuators at virtually every anatomical joint implies that rules must exist for defining how muscles will share the work load. Ref. [16] provides an extensive summary of ongoing research in muscle load sharing at the University of Wisconsin at Madison. An understanding of muscle load sharing is needed to explain why certain design options or operational scenarios have the potential of causing machine induced discomfort, fatigue, pain, or trauma.

### **PREDICTION OF HUMAN MOTION**

One fundamental difference between repetitively testing human subjects and repetitively testing mathematical models is that the human's response is nonrepeatable. The modeling goal for the prediction of human performance can therefore only be that the predicted motion be physically reasonable. Predictions and reasonable variations around them should be viewed as defining an envelop of possible human response. With this goal in mind simplified motion prediction algorithms can justifiably be introduced into a human motion prediction capability. The program JACK [17] has several unique features that make it ideal for the MMIDAP application. Figure positioning by multiple constraints is a capability that allows users to specify trajectories at several body fixed points (hand, feet, torso) and to then have motion trajectories for all other points predicted. Strength guided motion is a capability that uses human strength and comfort data for the motion prediction process. JACK uses a blend of kinematic, dynamic and biomechanical information when planning and executing a path. The task only needs to be described by a starting point, ending position, and external loads such as gravity and weights to be transported.

### **MAN/MACHINE DYNAMIC INTERACTION, ITERATIVE REFINEMENT**

The output of the program JACK is animated human system response. As for any engineering analysis study, the physical realizability of predicted response must always be checked. This is done by viewing animated response and resultant joint behavior. Weighting factors within the JACK program allow user's to tune predictions to bring them into the realm of physical realizability for the particular population group under study (old, young, normal, handicapped, etc.) As a further check JACK's predicted joint response information is used as input to the program Order N DISCOS. This program includes a detailed dynamics model for the machine and the operator's musculoskeletal system. The associated equations of motion for the multibody model are exact, relative to the laws of Newtonian mechanics. Order N DISCOS's inverse dynamics capability is then used to obtain a refined prediction for resultant joint loading. Differences between JACK and DISCOS resultant load predictions stem from the simplifying assumptions within JACK's motion prediction algorithms and man/machine interaction dynamics. The output of Order N DISCOS becomes input to the program defined in [16] to predict muscle load sharing. If resultant joint loading and muscle load sharing predictions are acceptable then motion and load prediction information is used as input to the human performance database at the HPI. The output of this step provides another assessment of physical realizability. If results violate physical realizability JACK weighting factors can be adjusted and the process

repeated.

The iterative refinement process is used until the successive approximations strategy converges to acceptable results. The predictions will either confirm that man/machine interaction is acceptable or that some human performance parameters exceed databased norms. Machine design changes can be refined until acceptable performance measures are achieved for the machine operator's population group.

#### **TARGET NASA SPINOFF APPLICATION**

GSFC's Office of Commercial Programs is strongly supportive of efforts to transfer aerospace developed technology. The following technology transfer initiatives are currently being pursued:

- o The underlying technology of real-time man in the control loop simulation is the enabling technology needed to build simulators for the ergonomic design of a man/machine interface. These facilities will support physiological and cognitive research related to human operation of mechanical systems with an emphasis on automotive vehicles. Issues associated with operator performance under the influence of alcohol, drugs, sleep denied fatigue, etc. can all be investigated in a simulator without the danger of life threatening injury to the test subject.
- o The ability to account for operator population group in the design process provides an ability to determine if a handicapped person has the physical and cognitive resources necessary to operate a particular machine. This capability not only opens the door to the job placement for the handicapped problem, but also allows machine designers to identify and remove design features that inhibit operation by the handicapped.
- o Exercise equipment can be viewed as human operated machines. Since operators tend to operate these near the limit of their physical resources, it is important that they be carefully designed. They must exercise particular muscle groups while not causing physical injury to other muscle groups or associated joint complexes.
- o Bone and muscle atrophy is a major problem for space missions. If the body senses that either bone or muscle is not needed it is absorbed by the body. To counteract this problem an exercise program must be designed to stress both bone and muscle. Proposed exercise systems can be evaluated relative to each astronaut and alternative exercise scenarios can be compared via quantifiable performance measures.
- o Physical therapy is a slow process that is difficult to measure in a quantifiable sense. Techniques used to validate MMIDAP capability will be directly applicable to the measurement of rehabilitation progress. The MMIDAP capability itself can be used to set exertion limits for recovering patients and machine operators.
- o Rehabilitation engineers develop a wide range of devices and prosthetics so the injured and disabled can recover lost performance and prevent further injury. Both active and passive prosthetic devices can be viewed as human operated machines that need MMIDAP design analysis.
- o Functional neural stimulation (FNS) is currently being used to stimulate both upper and lower extremity muscles of spinal cord injured patients. The objective of FNS research is to design a control system for the coordinated electric stimulation of those muscle groups needed to provide a standing, walking and grasping capability. Patient safety considerations demand that proposed FNS control strategies be evaluated via computational analysis before they are administered. An ability to measure and quantify patient progress is another need. The MMIDAP project will support both needs.

- o Occupational and sports related trauma is a serious problem when workers or athletes carry out strenuous repetitive actions or intentionally shock their musculoskeletal system; eq. carpal tunnel syndrome, white finger, and tennis elbow. A better understanding of how trauma causing forces and moments develop is needed to suggest less stressful body motion strategies.
- o Biochemists are attempting to better understand how drugs work within the human body. This quest leads to the need to model the molecular dynamics of protein enzyme reactions. The program Order N DISCOS is currently being beta site tested for this application at Dupont with the assistance of Cambridge Research. Attempts to use the program NBOD for this application are outlined in [18].

## SUMMARY

This paper has reviewed all essential existing capability needed to create a man/machine interaction dynamics and performance (MMIDAP) capability. The multibody system dynamics software program Order N DISCOS will be used for machine and musculo-skeletal dynamics modeling, the program JACK will be used for estimating and animating whole body human response to given loading situations and motion constraints,, the basic elements of performance (BEP) task decomposition methodologies associated with the Human Performance Institute's BEP database will be used for performance assessment, and techniques for resolving the statically indeterminant muscular load sharing problem will be used for a detailed understanding of potential musculotendon or ligamentous fatigue, pain, discomfort, and trauma. The envisioned capability is to be used for mechanical system design, human performance assessment, extrapolation of man/machine interaction test data, biomedical engineering, and soft prototyping within a concurrent engineering (CE) system.

## REFERENCES

1. Frisch, H.P. "NBOD & DISCOS - Dynamic Interaction Simulation of Controls and Structure," Multibody Systems Handbook, Werner Schiehlen (editor), Springer Verlag, pp 145-160, 1990
2. NSF/TACOM/NASA Industry/University Cooperative Research Center for Simulation and Design Optimization of Mechanical Systems, Participant Notebook, College of Engineering, The University of Iowa, Iowa City, Iowa 52242, phone (319) 335-5726, June 1989.
3. Frisch, H.P. "A Man/Machine Interaction Dynamics and Performance (MMIDAP) Analysis Capability," Second Annual Symposium on Mechanical System design in a Concurrent Engineering Environment, The University of Iowa, Oct 30-31, 1990.
4. Turner, J.D., "Integrated Ergonomic System Software Development," SBIR Final Report. Copy available from H.P. Frisch, Code 714.1 NASA/GSFC, Greenbelt, MD 20771 or J.D. Turner, Cambridge Research, 1033 Mass. Ave., Cambridge, MA 02133, phone (617) 354-1522.
5. Kroemer, K.H.E., et al. (editors), "Ergonomic Models of Anthropometry, Human Biomechanics and Operator Equipment Interfaces," Proceedings, Workshop on Integrated Ergonomic Modeling, 1988.
6. Fisher W.F. and Price C.R., "Space Station Freedom External Maintenance Task Team Final Report," NASA/JSC, July 1990.
7. "Electronic, Imaging," Report of the Board of Regents, National Library of Medicine Long Range Plan, NIH Publication Number 90-2197, April 1990.
8. Andrews J.G., Frisch, H.P., Kondraske, G.V., Mansour, J.M., Spitzer V., and Whitlock, D., "Plan for Integrated Biomechanics Analysis and Modeling: Tools and Databases," Final report of Whole Body Digital Mapping Project, 1990 (in preparation). Copy to be available from H.P. Frisch, Code 714.1, NASA/GSFC, Greenbelt, MD 20771, phone (301) 286-8730.

9. Kondraske, G.V. "Quantitative Measurement and Assessment of Performance," Chapter 6 of book *Rehabilitation Engineering*; Smith, R.V. and Leslie, J.H. (editors), CRC Press, 1990.
10. Stoner, J.W., et al., "Introduction to the Iowa Driving Simulator and Simulation Research Program," University of Iowa, Center for Simulation and Design Optimization of Mechanical Systems, Technical Report R-86.
11. Haug, E.J. et al., "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator," US Department of Transportation, DOT HS-807-596, 1990.
12. Schiehlen W. (Editor), "Multibody Systems Handbook," Springer Verlag, 1990.
13. Chun, H.M, Turner J.D. and Frisch, H.P., Order (N) DISCOS for Multibody Systems with Gear Reduction," Guidance and Control Conference, Portland Oregon, Aug 20-22, 1990.
14. Kim, S.S. et al., "New General Purpose Dynamics Simulation CODE (NGDC)," The University of Iowa, Center for Simulation and Design optimization of Mechanical Systems, 1990.
15. Winters, J.M. and Woo, S.L. (editors), "Multiple Muscle Systems Biomechanics and Movement Organization," Springer Verlag, 1990
16. Seireg, A. and Arvikar, R., "Biomechanical Analysis of the Musculoskeletal Structure for Medicine and Sports," Hemisphere Publishing Corporation, 1989.
17. Badler, N.L. "Human Factors Simulation Research at the University of Pennsylvania," Proceedings of the AFHRL Workshop on Human-Centered Design Technology for Maintainability," Air Force Human Resources Laboratory, Wright Patterson Air Force Base, Dayton Ohio, Sept 12-13, 1990.
18. "Drug Research," pp 83-84, SPINOFF, NASA Office of Commercial Programs, Technology Utilization Division, August 1989.

N91-23061

## TELEROBOTIC ELECTRONIC MATERIALS PROCESSING EXPERIMENT

Stanford Ollendorf  
Chief, Office of Telerobotic Engineering  
NASA Goddard Space Flight Center

### BACKGROUND

The NASA GSFC Office of Commercial Programs (OCP) is currently sponsoring industry and institutional research programs that will utilize GSFC developed robot technologies to facilitate space-based production of microelectronic materials (semiconductors) and devices. The microgravity space environment offers many advantages for microelectronic processing. In the absence of gravity, thermal convection, buoyancy and sedimentation are eliminated. Microgravity also enables containerless material processing, thereby reducing the likelihood of semiconductor contamination. A high vacuum controlled atmosphere, cleanliness and isolation of hazardous materials are other space processing advantages. The objective of the OCP and NASA Headquarters Space Commercialization Office, Code C, is to provide a broad program of technological research and development in support of space-based microelectronic materials processing applications and science programs by matching GSFC capabilities to industry and institutional needs.

#### Robotics for Commercial Microelectronics Processes in Space Workshop

The GSFC OCP hosted the "Robotics for Commercial Microelectronics Processes in Space Workshop" in December 1987. This was the beginning of a four Phase approach that will lead to a test payload for a space-based processing facility. The objective of the Phase I workshop was to bring together NASA, university, and industry researchers, as well as Space Shuttle carrier personnel, to discuss materials processing, robotics, and commercial automation. Discussed in the workshop was the utilization of robotics and automation to improve microelectronic processing productivity. Phase II of the program was initiated to investigate processes that included bulk crystal growth, wafer manufacturing and device fabrication. A "Telerobotics for Commercial Microelectronics Processes in Space" study report concluded that the utilization of automation and telerobotic technologies was far more cost effective and less hazardous than using a human operator. [Ref. 1]

It has been proposed that Phase III utilize the unique NASA GSFC Development, Integration and Test Facility (DITFAC) testbeds, shown in Figure 1, to identify key features for the material processes and demonstrate robot and automation capabilities. The DITFAC currently supporting the Flight Telerobotic Servicer (FTS) program incorporates three primary testbeds. These testbeds are being used by the GSFC robotics team to create, test and evaluate new and evolving robotic technologies. The testbeds include the FTS Functional Simulator, Engineering Testbed and the Graphics Simulator. These testbeds are being used to evaluate Space Station Freedom mockups, perform robot simulations, research issues such as robot control algorithms and perform graphic kinematic simulations of various robot tasks. [Ref. 2]

The Phase IV proposal represents the development of a Telerobotic Materials Processing Facility (TRMPX). TRMPX is a Shuttle-launched, materials processing test payload using a Get Away Special (GAS) can. The test payload is necessary because key features of the facility cannot be fully tested on earth and will require exposure to the unique environment of space. [Ref. 3]

ORIGINAL PAGE IS  
OF POOR QUALITY

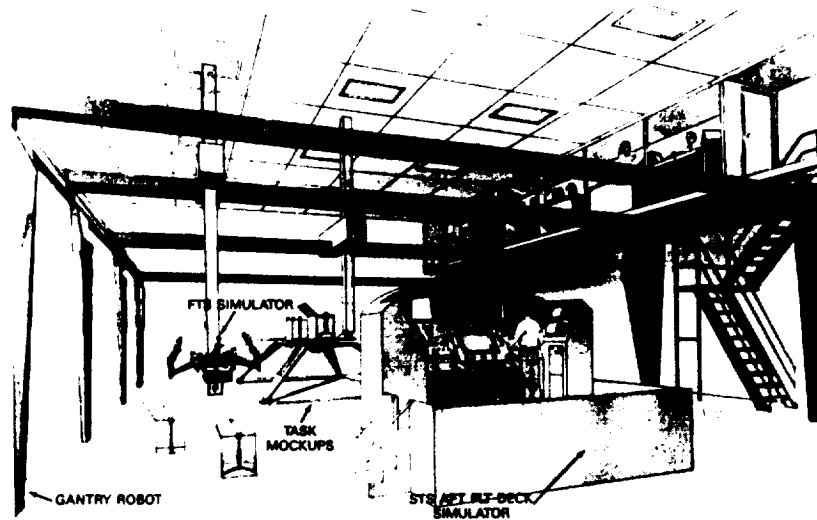


Figure 1. Development, Integration and Test Facility (DITFAC)

#### TELEROBOTIC MATERIALS PROCESSING EXPERIMENT (TRMPX)

The objectives of the GSFC managed and OCP sponsored TRMPX program are to define, develop and demonstrate an automated materials processing capability under realistic flight conditions. GSFC has proposed to accomplish this using small test payloads launched on the Shuttle as a Hitchhiker attached GAS payload experiment. TRMPX will fit inside the 19-3/4" diameter x 28-1/4" length GAS can interior. The Hitchhiker avionics and electronics interface will provide the necessary power and data handling to operate TRMPX. The process of utilizing the GSFC laboratory to develop and demonstrate robot capabilities, and the knowledge accrued from this experiment, reduce the risk in developing a space-based materials processing facility. A concept drawing of a proposed TRMPX is shown in Figure 2. [Ref. 4]

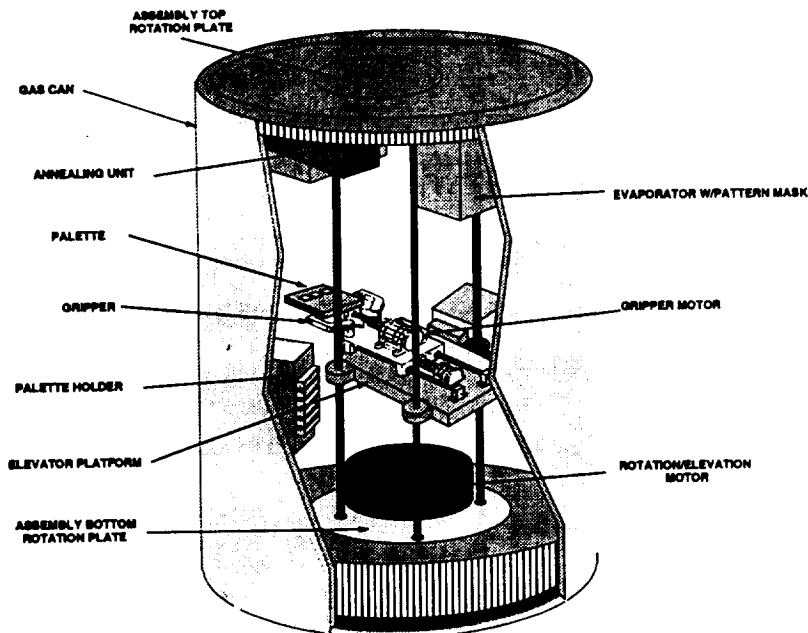


Figure 2. Telerobotic Materials Processing Experiment (TRMPX) Concept

### Materials Processing Experiment

TRMPX will provide the capability to test the production processes that are dependent on microgravity. The processes proposed for testing include:

- the annealing of amorphous silicon to increase grain size for more efficient solar cells,
- thin film deposition to demonstrate the potential of fabricating solar cells in orbit, and
- the annealing of radiation damaged solar cells.

### Robot Handling Equipment

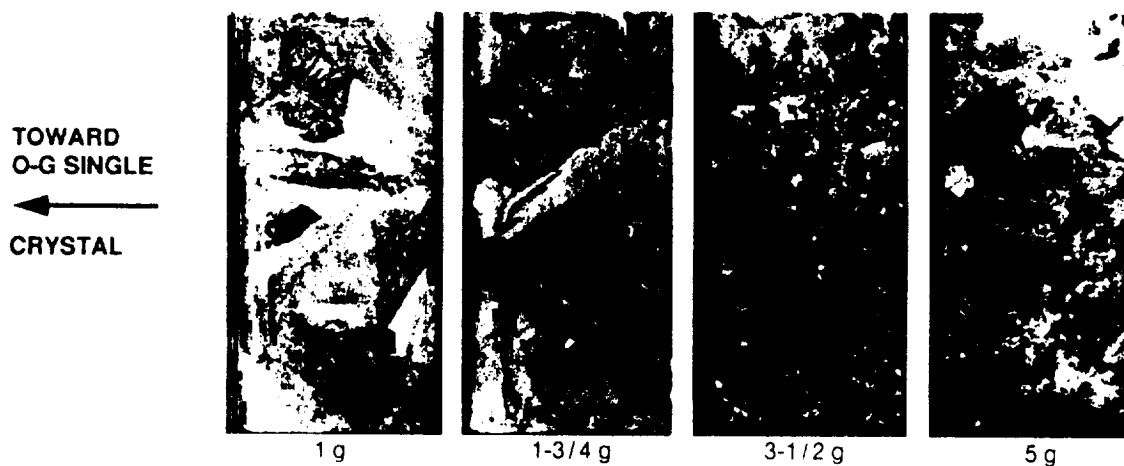
TRMPX robot-process testing will include:

- three axis robot manipulation,
- the gripper/grasping and robot manipulation of palette with sample materials, and
- robot control using autonomy as well as teleoperation from the ground.

### Mission Characteristics

The proposed materials processing mission for the initial TRMPX is the post fabrication recrystallization annealing of poly and amorphous Silicon (Si) to increase microcrystal grain size for the production of high efficiency solar cells based on inexpensive materials.

Figure 3 shows the effects on grain size for metal crystals grown in a centrifuge. It is theorized from this data that one can approach a single crystal configuration and therefore improve solar cell efficiency by processing in a microgravity environment.



### GRAIN VS. G-FORCE

Figure 3. Micro Gravity Effects

ORIGINAL PAGE IS  
OF POOR QUALITY



The robot interaction would include:

- substrate exchange from holding carriages to annealing area and back again,
- in-flight analysis and robot control to adjust the annealing cycles during the course of the experiment.

Robot tasks will further evolve as the experiment design is solidified.

The definition of TRMPX mission and robot design requirements will be completed during the second quarter of FY 91. A prototype Flight TRMPX will be available during the third quarter of FY 92 with a launch by the first quarter FY 93. These time lines assume continued funding support and STS availability.

#### Centers for Commercial Development of Space - CCDS

The TRMPX program utilizes in-house engineering support as well as contractors. Commercial Centers for Development in Space (CCDS) was established to provide for commercial involvement in requirements definition. The CCDS consists of universities and industry which conduct research and development of technologies with applicability to space-based materials processing. Clarkson College and Batelle Corporation are researching materials such as zeolite ampoules. They also plan to flight test GSFC silicon samples on-board a KC-135 aircraft. Flying at a high altitude, a free falling KC-135 can duplicate a weightless environment for short durations. Similar samples may be supplied to GSFC for the TRMPX experiment. The University of Wisconsin Center for Space Automation and Robotics (WCSAR) has been established to conceive, demonstrate and simulate space and terrestrial commercialization technology. The WCSAR has been involved in robot automation and sensor development and is considering joining the TRMPX team.

### TELEROBOTICS MATERIALS PROCESSING FACILITY (TRMPF)

#### Goals and Objectives

GSFC engineers have begun to define requirements and concepts for a free-flyer telerobotic materials processing facility (TRMPF). The objective of the TRMPF program is to evolve TRMPX to a fully self contained free flyer materials processing facility. TRMPF will utilize enhanced materials processing and robotic capabilities used on TRMPX.

#### Mission Scenario

The TRMPF facility, shown in Figure 4, will be approximately 30" diameter and 76" long. Initially, TRMPF would be an attached payload on the Shuttle evolving to a permanent facility on Space Station Freedom or a free flying spacecraft. The geometry and size of TRMPF make it a good candidate for launch on an expendable launch vehicle (ELV). The facility can be serviced by the Shuttle or an ELV with re-entry capability. Both would supply new materials for processing and remove processed samples and return to earth as shown in Figure 5. Another concept being considered is to launch multiple facilities that would be attached to each other by using an autonomous robot walker to form a TRMPF train as shown in Figure 6. Facilities that have completed their missions will be detached and returned to earth.

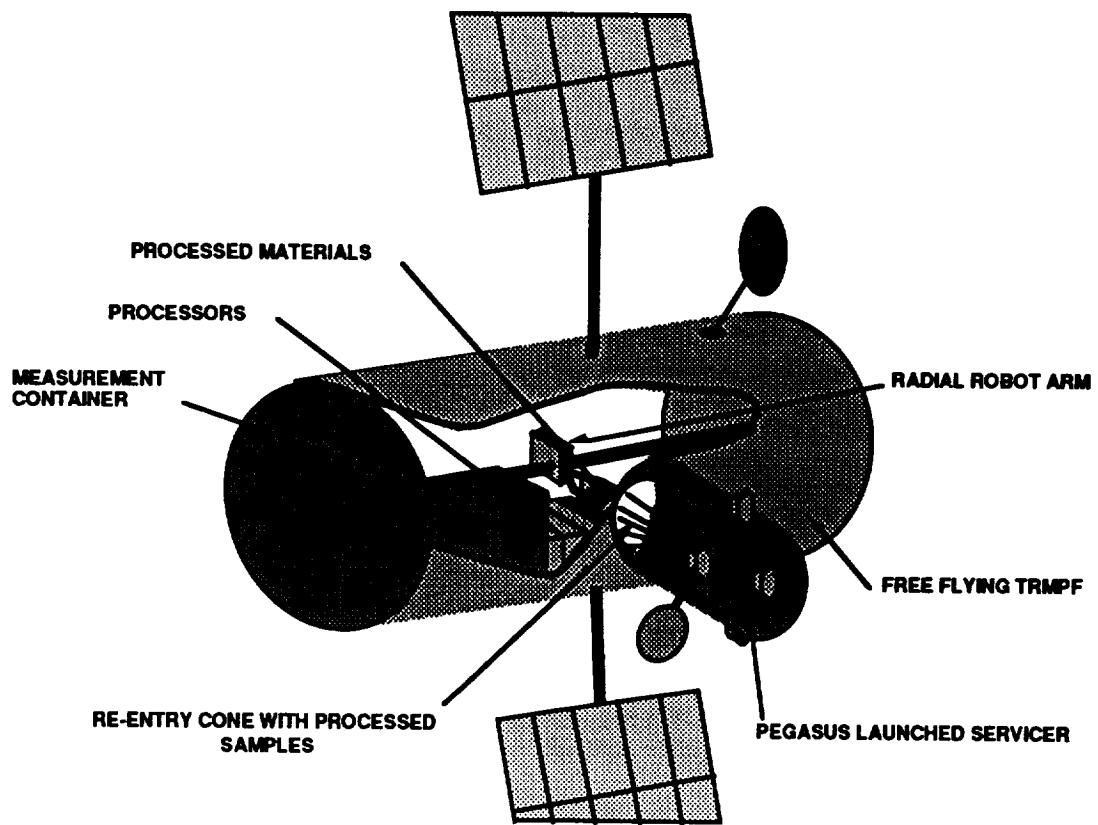


Figure 4. TRMPF Concept

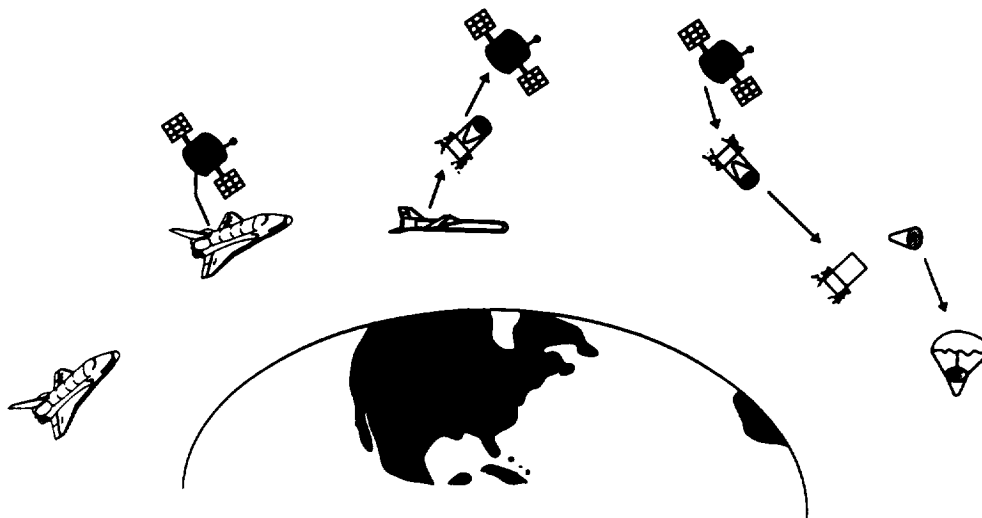


Figure 5. TRMPF Servicing

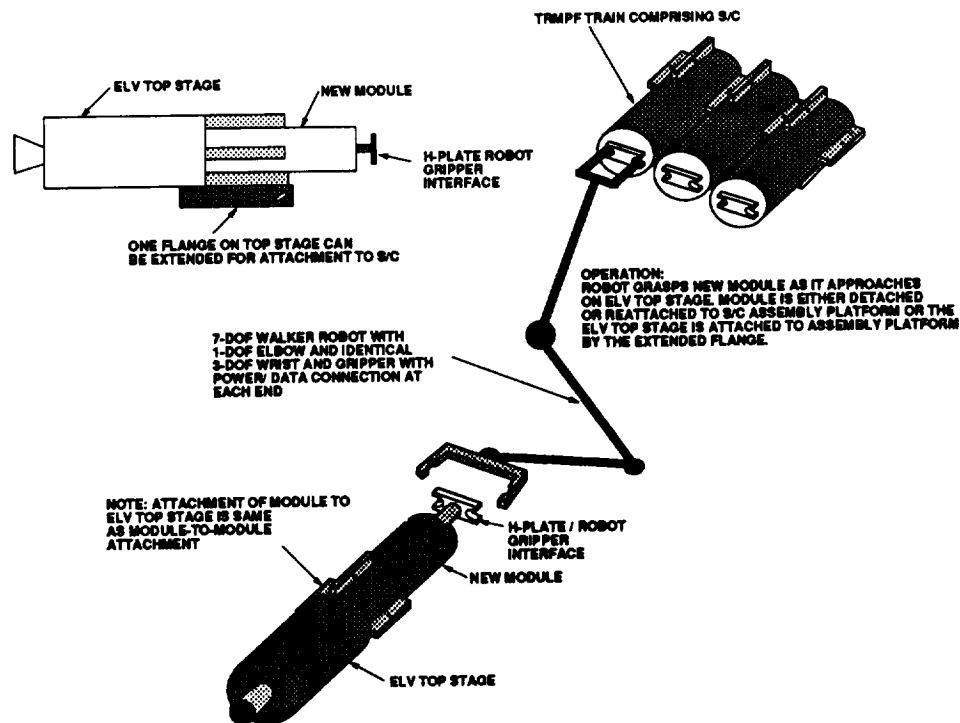


Figure 6. TRMPF Train

## SUMMARY AND CONCLUSIONS

The Office of Commercial Programs (OCP) working in conjunction with NASA GSFC engineers is supporting research efforts in robot technology and microelectronics materials processing that will provide many spinoffs for science and industry. The environment of space provides a number of potentially significant advantages in regard to commercial microelectronic processes. To provide for an environment by which these emerging technologies can be tested and demonstrated, GSFC has developed and implemented a Development, Integration and Test Facility. The OCP is sponsoring universities and industry to participate in this tremendous opportunity to contribute to the development of microelectronic processing and robotic applications in space.

TRMPX is being designed as an attached Shuttle payload to demonstrate an automated materials processing capability under realistic flight conditions. TRMPF will evolve using the knowledge gained from TRMPX to form a stand-alone self-supporting free-flyer materials processing facility in space.

## REFERENCES:

1. "Telerobotics for Commercial Microelectronics Processes in Space." NASA GSFC, 1987.
2. Ollendorf, S.: "Robotics Program at NASA Goddard Space Flight Center." NASA GSFC, 1989.
3. "Telerobotic Electronic Materials Processing Experiment (TREMPEX)." Systems Concept Review. NASA GSFC, December 21, 1989.
4. "Get Away Special (GAS) Small Self Contained Payloads - Experimenter Handbook." NASA GSFC, 1989.

## ADVANCED MECHANISMS FOR ROBOTICS

John M. Vranish

NASA/Goddard Space Flight Center  
Greenbelt, Maryland 20771

## INTRODUCTION

In recent years NASA has launched an extensive effort in robotics for space applications; the idea being to assist and augment manned space flight and, perhaps equally important, to give a major boost to U.S. industrial competitiveness by transferring the technology into the private sector. In this effort, research on mechanisms, actuators and motors and their associated sensors and controls, has traditionally been sacrificed on the altar of a rush for more and more "compute power", sophisticated graphical and animation packages and controls techniques. This, of course, results in lopsided progress; twenty first century computers and controls coexisting side-by-side with nineteenth century mechanisms, actuators and motors. But; robotics is a systems challenge and if one portion of the system is limited, the system is limited. A balanced approach to R&D is a must if real systems progress is to be made. The intent of this paper is to show what is being done at GSFC to bring about this balance and, as a by-product, the benefits and opportunities open to U.S. industry.

Accordingly, an overview of applied research and development at the Goddard Space Flight Center (GSFC) on mechanisms and the collision avoidance skin for robots is presented. First the work on robot end effectors is outlined, followed by a brief discussion on robot-friendly payload latching mechanisms and compliant joints. This, in turn, is followed by the collision avoidance/management skin and the GSFC research on magnetostrictive direct drive motors. Finally, a new project, the artificial muscle, is introduced. Each of the devices is described sufficiently to permit a basic understanding of its purpose, capabilities and operating fundamentals. In addition, the development status of each is reported along with descriptions of breadboards and prototypes and their test results. In each case, the implications for commercialisation is discussed. The chronology of the presentation will give a clear idea of both the evolution of the R&D in recent years and its likely direction in the future.

## I. END EFFECTORS

In this section, GSFC end effector R&D is discussed. The GSFC "Gripper-Nut Runner" (fig. 1) is described first, followed by a new emerging concept based on "Spline Screws".

The "Gripper-Nut Runner" has been under development at GSFC for more than three years and has evolved to the point shown in Fig.1. Since space is a "micro-g" environment, objects must be fastened to something (say, for example, Space Station) to prevent their drifting away. The function of the "Gripper-Nut Runner", therefore, is to grasp a dedicated interface attached to such an object using the gripper and to use the nut runner to loosen the fastener which fixes it to Space Station. This permits the robot to grasp objects and unfasten them and to move and re-attach them. This system has been proven in the GSFC robotics lab as well as several other NASA-affiliated activities around the country. It is rugged and durable and has repeatedly withstood forces at the finger tips on the order of 200 lbf. Of its component subsystems, three have commercial possibilities.

a. The "Split-Rail" Parallel Gripper [1]. This device (the gripper portion of the system shown in Fig. 1), was designed and patented by a NASA engineer, and has obvious commercial possibilities as a general purpose industrial gripper. This is a high performance wide throw parallel action gripper that uses a unique "split-rail" concept to make it simple, light, and compact (hence, inexpensive to manufacture). It is made primarily of anodized aluminium with straight machining cuts (no grinding) and rolls on cylindrical bearings throughout its stroke and so is strong, precise, responsive and will not jam under side loads.

b. Very Large Scale Integration (VLSI) gripper controller [2]. This state-of-the art gripper controller was developed as part of a NASA Small Business Innovative Research contract (SBIR). It is a dual-axis system that incorporates unique and innovative custom Application Specific Integrated Circuits (ASICs) to permit unusual performance in terms of filtering, precision and smoothness of motor control. It has a unique and compact power supply system to permit its being mounted on the gripper. Interfaces, hardware and software are kept serial and simple. This system is fully developed and functional. It has obvious commercial implications for any type of servo

joint. For example, it can readily be adapted, in a modular fashion, to form the basis for an entire robot control system.

c. Rolling Friction Fingers [3]. These simple and compact gripper fingers (Fig. 1) provide a vectored, rolling friction guidance and locking system between the robot gripper and the dedicated interface. Invented and designed in-house at GSFC, these fingers enable the gripper to acquire a dedicated interface of an object despite large initial misalignments and to guide the gripper to a seat and lock with the object, providing low friction and smooth operation throughout, despite the presence of large side forces. It permits the gripper to release the object under strong side loads, a very important safety hedge against the object being caught and jammed in the jaws. Also, because of its rolling action, it cannot scar or "burr" the object regardless of the forces. And, because of this rolling action, force feedback sensing is cleaner and the size of the motors required to drive the gripper screw can be reduced. Most of the GSFC grippers are equipped with this device. It has proven out in the laboratory over continued use; in one instance, for example, permitting a weak gripper motor to release an object under side loads of 100 lbf. and consistently making it easier for robots to acquire and grasp objects using both passive and/or active compliance. This should have a wide range of commercial applications wherever guidance and latching and locking of objects is required under conditions of large misalignments and opposing forces.

d. Wrist-Driven Auto Changer [4]. Invented and developed in-house at GSFC, this simple, compact and strong device provides a safe and reliable means for space robots to exchange end effectors. It is actuated by a unique simple camming action which occurs between mechanisms in a tool interface plate, a keying element in the tool storage holster and another mechanism attached to the robot as the robot stores or removes the tool. The entire system is passive with all action being initiated by the robot actuators and controls. Hence, many of the sensors of the robot joints can be used (to include the encoders or resolvers and the wrist 6 vector force feedback sensor) to aid in monitoring the status of and controlling the Auto Changer. Also, many of the interfaces to robot control are already inherently taken care of. The device is inherently safe and reliable because one of the prime sources of Auto Change mishap, inadvertent release, is impossible. It has been tested extensively in the GSFC robot lab and has met or exceeded every expectation. It has even been operated in a teleoperator mode without force feedback-a very difficult feat. This should have a limited; but significant niche' market in the commercial world.

GSFC is in the process of developing an entirely new approach to end effectors based on a simple straight-forward concept called "Spline Screws" [5]. This concept was invented and developed in-house at GSFC and represents an entire space fastening strategy of which end effectors is but one portion. We will develop the explanation of this concept by a simple example (fig. 3). We assume that the robot wrist has a dual roll capability about a common center. The inner roll terminates in a splined screw driver and the outer roll provides the means to rotate the object being grasped. This constitutes our end effector. The object to be moved is pierced by a screw (typically 0.5 in. dia.) which is splined so as to mate with the splined screw driver on the one side and the fixture to which the object will be attached on the other. The place of attachment has a small rotating fixture which is splined to cooperate with the piercing screw of the object on the side away from the gripper. We will begin by assuming the object is fastened to an attachment point. The robot would position the end effector over the splines of the screw piercing the object. The robot would guide and seat the inner roll splined screw driver and then guide and seat the outer roll torque tabs into slots in the top of the object. The splined screw driver would be turned clockwise. The splines of the screw driver would mesh and lock with those of the object screw, the object screw would turn with the rotating fixture. In the process, the object screw would translate towards the attachment fixture. This would unlock it from the fixture and lock the object to the end effector. The end effector would be free to leave with the object and maneuver it pending attachment to another fixture. Once aligned with and seated on this new fixture, the end effector would be turned counter clockwise and the object screwed off the end effector and onto the new fixture. The object would be either attached to the end effector or the attachment fixture or both at all times.

The splines permit the end effector screw driver to capture the object screw yet are sufficiently coarse that cross-threading between the two would be impossible. The same would be true of the interface between the object screw and the attachment fixture. On the other hand, the object screw would be captive in the object so it could have a very fine thread with a short lead and be lubricated to give great holding forces with modest amounts of torque (approx. 600 lbf from 8 ft-lbf torque typical). The system would not backdrive so brakes would not be necessary on the screw driver and an attached object could survive launch forces. It is also clear that foot prints on the object and the attachment fixture would be very small (on the order of 1.25 in dia.) as would the end effectors. The system would be basic, simple, strong and very reliable. It would also be extremely versatile.

To demonstrate the versatility of the "Spline Screw" technique the example of an Auto Change is given. In this case we attach the screw to a common tool interface by means of slightly compliant wavy springs. But where

the nut was the pierced object in Fig. 2, it is an electrical connector in the Auto Change which is partially trapped so that it can move in translation only with respect to the common tool interface. Thus, as the screw is turned, the electrical connector nut would translate upwards, inserting the pins into the end effector electrical connector and releasing the Auto Change and its tool from its holster. At this point, the connector nut would be stopped by the upper portion of the Auto Change. But; the screw would still be turning so it, in turn, would translate downwards towards the tool, compressing the lower wavy spring washer in the process. In doing so, however, The Auto Change and end effector would seat and lock together. The process of storing tools would be the reverse. The screw would be turned counterclockwise, the nut would translate down until the tool locked to the holster. In the process the tool and electrical pins would separate from the end effector. Torques required would be modest (approx. 8 ft-lbf.) and the system would seat and lock or store with authority and certainty. And, of course, the dimensions and weights would be minimal.

## II. ROBOT-FRIENDLY PAYLOAD LATCHING MECHANISMS.

It appears the most efficient robot-friendly payload latching mechanism to date would simply use the "Spline Screw" approach in yet another configuration. The design is closely related to the Auto Change described above; with a few minor additions.

## III. COMPLIANT JOINTS.

Compliance is a critical component in the interaction between the robot and the object being grasped. Without it, a robot cannot capture, seat and lock an object. This almost always involves using passive (typically springs) and active (robot movement through sensor information) compliance in a cooperative manner. A GSFC engineer has pioneered numerous inventions in the area of passive compliance using a novel "Compliant Cable" [6] approach. With this approach cables can be wound and arranged in light weight, strong, and compact structures to provide vectored compliance. That is, the spring constant can be tailored independently in each of the six vectorial directions. The compliant cables, themselves, are composed of several strands wound around each other so they have a spring effect (coupled with impressive tensile strength); but with sufficient friction between the strands that the spring oscillations are damped out. This simple concept has been used as the basis for many devices in government and in industry of which only a portion would come under the heading of robotics. For example, these cables have proven to be excellent shock absorbers and vibration isolators for use in space and are also used extensively in the GSFC Robotics Lab. Commercial devices, based on this principle are already in extensive use. This is an important niche market.

## IV. COLLISION AVOIDANCE/MANAGEMENT SKIN.

Safety is a prime concern for robots operating in space; particularly when they are operating near humans and/or space structure. Thus, NASA is developing a collision avoidance/management skin wrapped around the robot arms. This skin will consist of an array of sensors each of which is called a "Capaciflector" [7] (or capacitive reflector). Invented and developed at GSFC, this technique enables a capacitive sensor to be mounted in the immediate vicinity of the grounded robot arm and still "see" out to ranges an order of magnitude further than previously reported. For example, we routinely demonstrate picking up a human hand or a four in. dia. aluminium cylinder at ranges in excess of one foot using a 0.25 in. wide, four in. long strip of copper tape as a capacitive sensing element with an operating frequency of 20 khz and a potential of 10 volts. The previous state-of-the-art range of such a sensor is approximately one inch. Normally the electric field of a capacitive sensor couples both back into the ground plane and out towards an approaching object. The less the stand-off from the ground plane, the more the coupling into that ground plane and the less the coupling towards the approaching object. In the "Capaciflector" a shield is interjected between the ground plane and the sensor. This shield is driven at the same frequency and is in phase with the sensor. It is also at the same potential. However, it is electrically isolated from the tuning portion of the oscillator. Thus the sensor couples with an approaching object and changes the frequency of the oscillator (detection by standard fm techniques), whereas, the shield follows that oscillator frequency; but any coupling it may do to ground or the object does not effect operating frequency. The result is that for the sensor to couple to ground, its electric field must go around the shield to reach ground. Thus we have the effects of a very large stand-off. The sensors and shield, however, may be part of a very thin flexible printed circuit board mounted directly on the ground plane. The commercial prospects for such a sensor are, of course, very significant. It will, no doubt, one day be a standard feature of many of the millions of industrial capacitive sensors at work throughout industry.

## V. MAGNETOSTRICTIVE DIRECT DRIVE MOTORS.

Current robot motors are very high speed; but have weak torque compensated by using a transmission with extensive gearing. Since safety brakes are also required in these joints, these brakes must be located to act on the motor itself or the drive shaft on the motor side of the transmission to give them sufficient holding leverage. All these additions, compensations and restrictions lead to complications, lower reliability and controls problems. The magnetostrictive motor project addresses these concerns by developing a device that has outstanding torque density and is self-braking with the power off. This permits the power to be taken directly off the drive shaft, eliminating brakes and transmissions. The magnetostrictive phenomenon using the material Terfenol-D shows promise because it generates impressive forces ( $> 4$  ksi) and has excellent frequency response (6 khz for 0.25 in. dia. rod). However, it also has two significant drawbacks, it has a very short stroke (0.001 in./in.) and low magnetic permeability (5) [8]. These present formidable engineering challenges.

Two engineering approaches are pursued (Fig. 3); type A using the classical "inch worm" approach and type B using an original (more promising) approach based on a roller locking technique. A proof-of-principle type A device has been successfully tested. It produced 9 ft-lbf stall torque (a record for an electric motor of this size), had a 800 microradian step size (outstanding for precision control), consumed 600 watts power and had a no-load speed of 0.5 rpm. Sound generated by the pounding of the clamping rods was surprisingly modest. The device was 10.25x4.50x4.25 in. and weighed 39 lbs [9]. The weight is not significant, nor is the low speed since no effort was made to control weight and limit the inertias of the moving parts at this stage of development. Also, small diameter rods were used so even this breadboard is fundamentally underpowered. Never-the-less it is clear that with development, this motor will become very competitive with torques on the order of 100 ft-lbf and no-load speeds near 20 rpm. We are now poised to begin work on prototype B, bringing it to the same level as A. It should exceed A in torque, speed and have outstanding efficiency.

From a commercialisation point of view, two more years of development will be required before these motors are ready. Ultimately, however, we are expecting a niche' motor which will significantly extend the state-of-the-art in applications requiring low speed, safe, high torque in a modest-sized package.

## VI. ARTIFICIAL MUSCLE.

It is clear that mechanisms, and hence robotics systems, are limited because we do not have a linear motor/actuator which can perform the functions of the basic muscle. Such a motor must both perform at the level of the human muscle (strength, compactness, linear stroke, frequency response, and controllability) and be able to consume fuel and produce power independent of an umbilical or large battery for expended periods of time. We are not seeking to reproduce the human muscle; only its performance. This project is being initiated at a highly qualified university under NASA/GSFC direction. It is realistic and results and products are, we feel, a near certainty. However, they are approximately three years away. The eventual commercial implications of this work will, of course, be very significant.

## SUMMARY AND CONCLUSIONS

Highlights of the NASA/GSFC program in advanced mechanisms and sensors have been presented. From this overview four things should be clear. 1. There is a tremendous amount of work to be done, from basic screws to futuristic muscles. There are crippling inadequacies in our present capabilities every where. 2. The existence of these inadequacies is distorting and hindering the progress of robotics in general. 3. NASA/GSFC is doing everything humanly possible to plug some of the more glaring holes as quickly as possible. 4. There are important commercialisation opportunities to be realized throughout.

## REFERENCES

1. Vranish, J.M. and Sharifi, Mohammad: The NASA/GSFC Split-Rail Parallel Gripper. Proc. 2nd European In-Orbit Operations Technology Symposium, Toulouse, France, September 12-14 , 1989.
2. Brown, J., NAVTROL Corp., Small Business Innovative Research Program-Controller demonstration at GSFC, July 27, 1990.
3. Vranish, J. M. : Rolling Friction Fingers. Research and Technology-GSFC Annual Report 1989, pp. 218-219.
4. Voellmer, G. Patent application entitled "Robotic Tool Change Mechanism", dtd. October 12, 1990.
5. Vranish, J. M. Invention Disclosure entitled "Spline Screw Fastening and End Effector System", dtd. October 22, 1990.
6. Kerley, J. Personal notes.
7. Vranish, J. M. and McConnell, R.L. Invention Disclosure entitled "Driven Shielding Capacitive Sensor ", dtd. August 20, 1990.
8. Terfenol-D notes, Publication of Edge Technologies, Inc. ETREMA Products Division Volume 3, #1, March 1990.
9. Glapa, S. Honey Robotics, Inc. Test Report dtd. October 10, 1990.



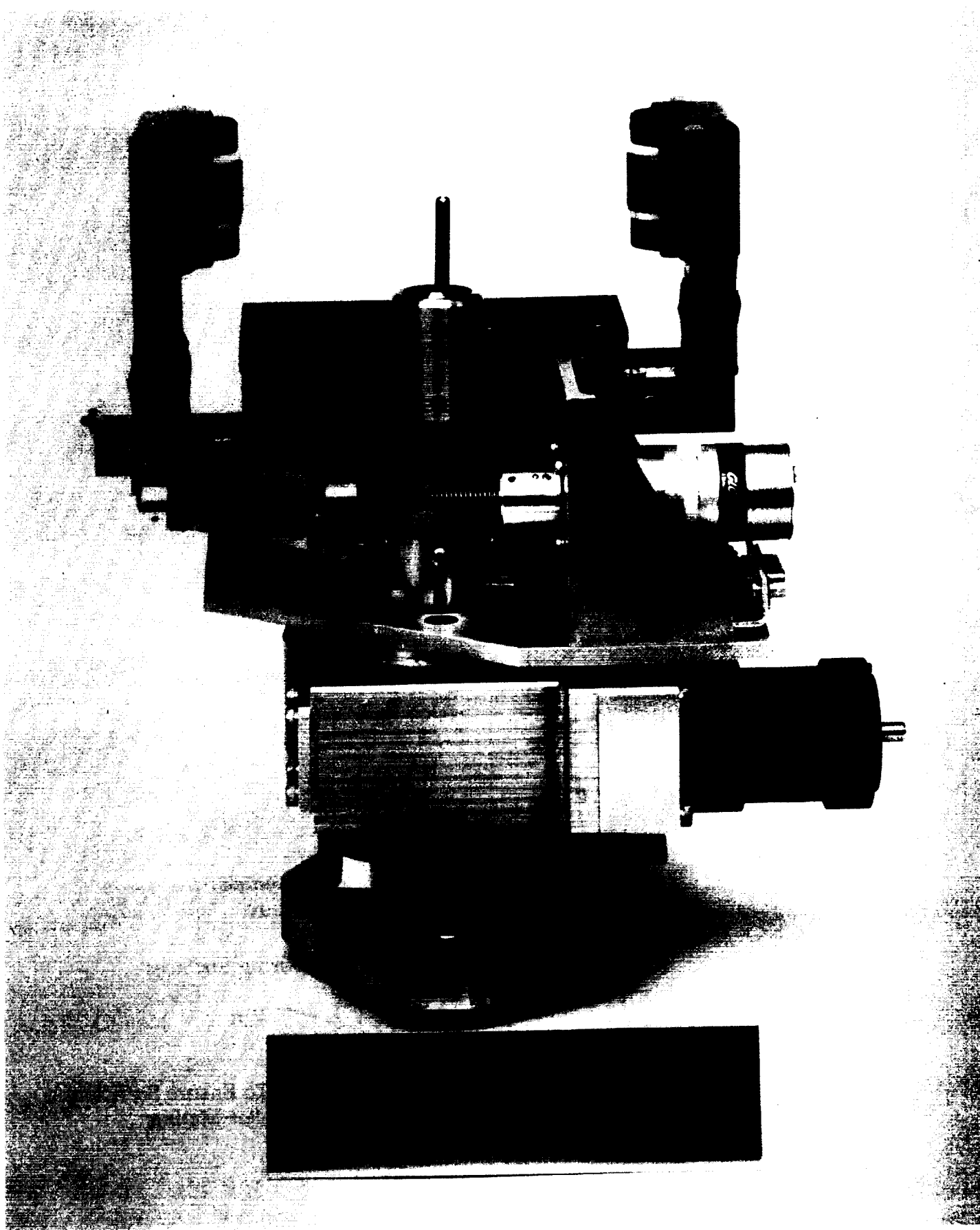


Fig. 1 THE GSFC "GRIPPER NUT RUNNER"

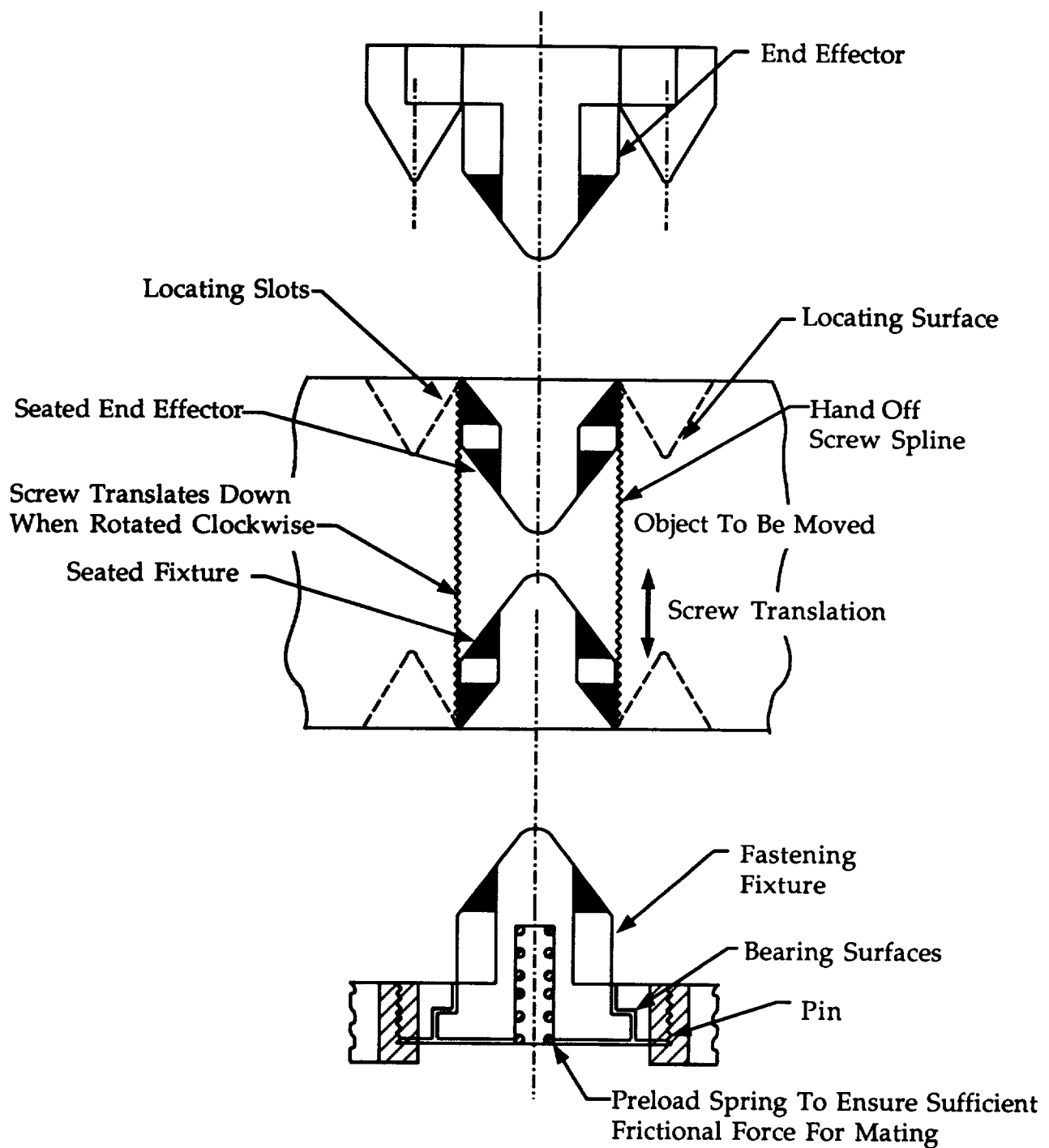
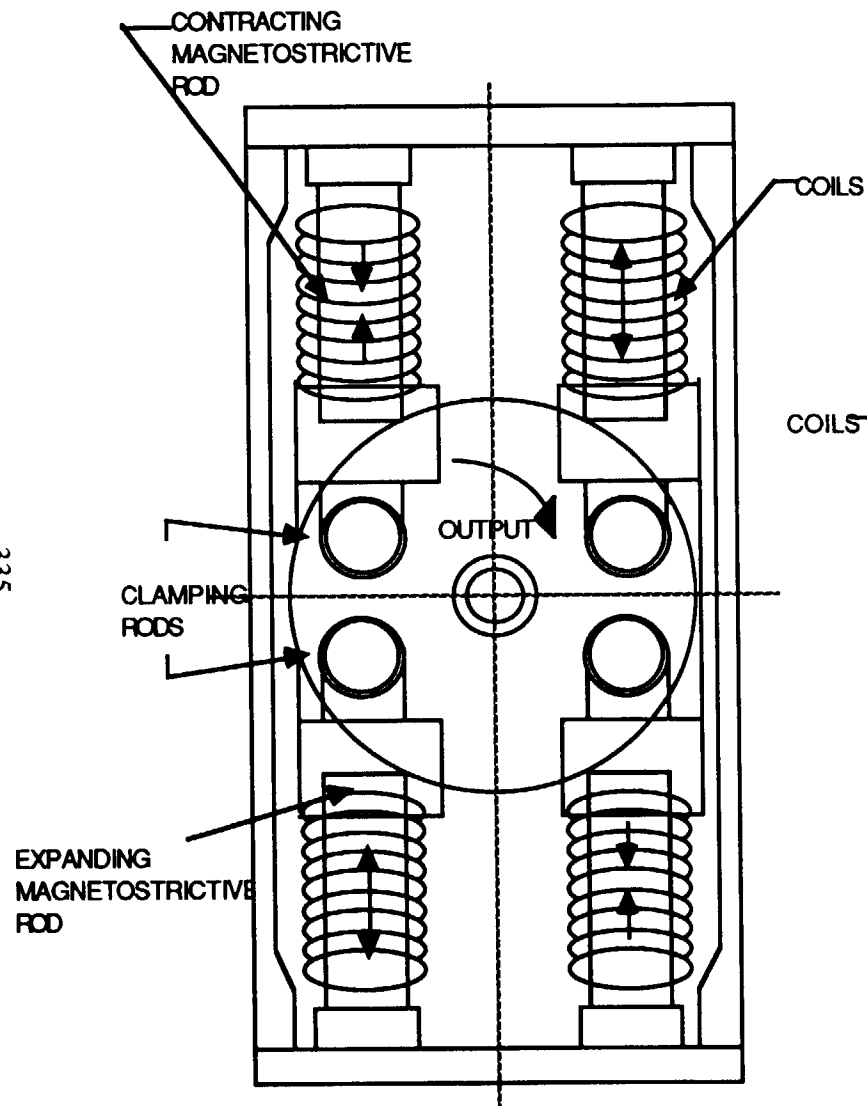


FIG. 2 FASTENING USING LOCKING SPLINES

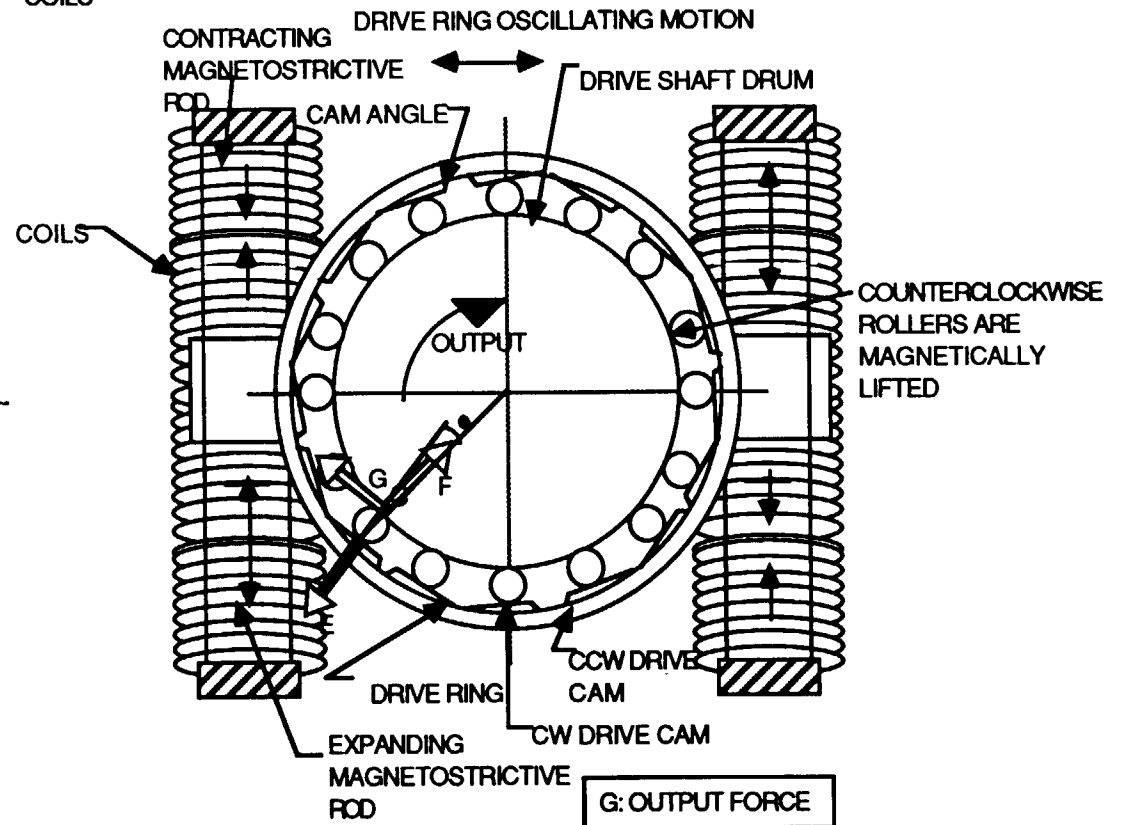
FIG. 3 MAGNETOSTRICTIVE DIRECT DRIVE MOTOR

COMPETING DESIGN CONCEPTS

335



TYPE A: MAGNETOSTRICTIVE DRIVE  
MAGNETOSTRICTIVE CLAMP



TYPE B: MAGNETOSTRICTIVE DRIVE  
ROLLER LOCKING CLAMP

N91-23063 !

## THE FLIGHT TELEROBOTIC SERVICER AND TECHNOLOGY TRANSFER

James F. Andary  
FTS Systems Manager  
Goddard Space Flight Center

Kayland Z. Bradford  
Commercial Applications Manager  
Martin Marietta Astronautics Group

### ABSTRACT

The Flight Telerobotic Servicer (FTS) Project at the Goddard Space Flight Center is developing an advanced telerobotic system to assist in and reduce crew extravehicular activity (EVA) for Space Station Freedom (SSF) [1]. The FTS will provide a telerobotic capability in the early phases of the SSF Program and will be employed for assembly, maintenance, and inspection applications. In mid-1989, the FTS Project entered the flight system design and implementation phase (Phase C/D) of development with the signing of the FTS prime contract with Martin Marietta Astronautics Group in Denver, Colorado. The basic FTS design is now established and can be reported on in some detail.

The current state of space technology and the general nature of the FTS tasks dictate that the FTS be designed with sophisticated teleoperational capabilities for its initial primary operating mode. However, technologies, such as advanced computer vision and autonomous planning techniques, would greatly enhance the FTS capabilities to perform autonomously in less structured work environments. Another objective of the FTS program is to accelerate technology transfer from research to U.S. industry.

### SYSTEM DESCRIPTION

The FTS flight system consists of the telerobot, two workstations -- one for the shuttle orbiter and one for Space Station Freedom -- and a facility for on-orbit storage, called the Storage Accommodation Equipment (SAE).

#### The Telerobot

The telerobot [figure 1] has two 7 degree-of-freedom (DOF) manipulators and an attachment, stabilization, and positioning subsystem (ASPS) or "leg" mounted on a compact body. The body contains internal electronic boxes that provide the power, data management, and data processing functions. The boxes are designed as orbital replaceable units (ORUs) that can be replaced by an astronaut on extravehicular activity (EVA) or another telerobot. Also mounted on the body are a camera positioning assembly with two head cameras and holsters for storing tools and end effectors. On each manipulator wrist actuator is mounted a camera for close-up viewing. At the end of the wrist roll actuator is a force-torque sensor for measuring forces and torques produced at the tool plate. Attached to the force-torque sensor is the end effector changeout mechanism, which will accommodate a variety of tools and end effectors.

#### Workstations

The shuttle and space station workstations will provide the operator with similar interfaces, including color video displays, text and graphics overlay capability, and two six-DOF force reflecting handcontrollers for teleoperation of the FTS manipulators. During operation, a sequence of events is displayed which the operator will use as a checklist. As commands are issued, displays provide status information and command menus. Anomalous events result in automatic caution and warning displays to the operator, along with fault diagnostics and recommended corrective action.

### Storage Accommodation Equipment

The FTS will be stored on orbit at the SAE which will be attached to the space station truss on the nadir facing side. Also stored in the SAE will be the telerobot's ORUs and tools. The SAE has power, data, and video interfaces with the SSF, and an EVA station to support EVA maintenance of the FTS.

The weight of the telerobot and workstation will be approximately 1500 pounds. The average on-orbit power is planned to be less than 1000 watts with a peak power of approximately 2000 watts. The system is designed for an indefinite on-orbit life through periodic maintenance.

### Operating Modes

The FTS can operate while attached to a worksite or while attached to a transport device [figure 2]. The way the FTS is attached to a worksite is with its ASPs. If it is at an "improved" worksite, the telerobot connects to the power, data and video systems through a worksite attachment fixture (WAF) which also gives the telerobot mechanical support. If it is at an "unimproved" worksite, the telerobot will receive its power, data and video utilities through an umbilical that will be connected to a nearby utility port. In this case the worksite only provides mechanical support. In the transporter-attached mode, the telerobot is attached to a shuttle or space station remote manipulator system (RMS) via a grapple fixture mounted on the back of the FTS body.

### Test Flights

Two test flights -- a Development Test Flight (DTF-1) and a Demonstration Test Flight (DTF-2) -- precede the deployment of the initial operational FTS system at first element launch (FEL) in 1995. DTF-1, scheduled for launch in 1992, will validate the performance of the FTS manipulator design in a zero-gravity environment [figure 3]. Data obtained will also be used to evaluate human-machine interfaces, assess servicing interfaces with work elements, and establish the correlation between simulation and analysis results and flight performance data. DTF-2, scheduled for launch in 1994, will validate the full servicing capabilities of the FTS. Following DTF-2, the DTF-2 flight hardware will be refurbished, updated as required, and installed at the Goddard Space Flight Center as an engineering test system to support operation and evolution of the FTS.

## **CURRENT FTS TECHNOLOGY**

Engineering and integrating the different telerobotic technologies into an operational system that combines teleoperation and autonomy and that meets stringent performance requirements is the challenge of the FTS project. The following paragraphs summarize the major elements of these technologies.

### Manipulator Technology

The FTS manipulator [figure 4] comprises harmonic drive actuators with brushless d.c. torque motors commutated by Hall effect sensors. The transmissions have 100:1 speed reduction and incorporate fail-safe brakes for safety purposes. Table 1 shows manipulator performance specifications. A force-torque sensor attached to the manipulator tool plate measures the forces and torques exerted at the worksite and provides feedback to force reflecting handcontrollers at the workstation. It also allows the telerobot to control the forces and torques generated by the manipulators using a position-based, impedance control algorithm. This compliance control also allows the manipulators to automatically compensate for misalignments. The operator may select position or resolved-rate control, or in special situations, single-joint control. Coordinated dual-arm control will be useful in handling large components on Space Station Freedom.

Three microprocessors are imbedded in the manipulator links to control the seven joints and the end effector. Flat conductor cables carry hundreds of electrical signals through the actuators and back to the central processor in the telerobot body where the bus controller routes the commands and signals between the workstation and the telerobot. Several types of signals, including a 1553 data bus and the video channel from the wrist camera mounted to the wrist roll actuator, are passed through the actuators.

## Control System Technology

The NASA/NBS Standard Reference Model (NASREM) control system architecture, which has been selected for FTS, allows teleoperation and autonomous operations. Its architecture is a three-legged hierarchy of computing modules (figure 5). The first leg plans and executes the decomposition of high level goals into low level actions. The second leg remembers, estimates, predicts, and evaluates the state of the world surrounding the telerobot. The third leg recognizes patterns, detects events, and filters and integrates sensory information. The sensory system compares the world model predictions with observations. Each level in the hierarchy has a specific function, and receives commands from the next higher level. The first four levels of NASREM will be implemented in the FTS system. Level 4 decomposes tasks into sequences of movements. Level 3, the elementary move level, plans all aspects of the manipulation. The primitive level, or Level 2, generates the time sequence of desired state vectors to produce dynamic trajectories. The servo level, or Level 1, contains the servo control loops for the actuators.

The FTS flight software, coded in Ada, interprets the operator inputs and controls the telerobot's motion, monitors health and status, and ensures the safety of the FTS. The primary control loop operates at 50 Hz, which means that all the control computations and the data transfer from the handcontrollers to the manipulators back to the handcontrollers must be accomplished in 20 milliseconds, including data transmission delays.

## **FTS GROWTH TECHNOLOGY**

In the unstructured working environment of the FTS, current technology dictates the FTS operational mode be primarily teleoperation. However, technologies, such as artificial intelligence planning and advanced computer vision, are evolving from the research phase and could significantly enhance FTS capabilities to perform tasks autonomously in space. Therefore, the capability to evolve is designed into the FTS. In the meantime, it must use the operator for those tasks for which the human is better suited, such as vision processing and manipulator path planning [2].

The FTS project relies on a number of sources for the development of the growth technologies which will be needed for the FTS evolution. The primary sources of these advanced technologies are NASA's Office of Space Station (OSS) Advanced Development Program and the Telerobotic Development and Demonstration Program of the Office of Aeronautics and Space Technology (OAST). The FTS project works directly with these organizations to establish technology requirements. The most promising technologies will then be considered for incorporation into the FTS system as part of its planned evolution process.

### Preliminary Requirements

One example of the evolutionary planning that is being conducted by the FTS project is a recent study [3] that assessed the impact of future (1999) requirements on the FTS data management and processing system (DMPS). A set of preliminary requirements were generated based on an estimate of available state-of-the-art technology without regard to current flight qualification status and without power, weight, redundancy, and safety considerations. These preliminary requirements will be refined into a coherent, viable set that will be used to focus the technology development process. The following paragraphs summarize these preliminary requirements.

**Control.** The telerobot will be able to operate at NASREM level 4. Head cameras will automatically track specified objects. End effectors will incorporate "smart" 6-DOF micromanipulators. Other growth requirements include coordinated automation and teleoperation, control with time delay (e.g., ground control), increased speed, and control by implementing different manipulative algorithms. Improved collision avoidance techniques using geometric models and proximity sensor data will also be incorporated.

**Vision.** Requirements for presenting stereo vision to the operator and for color in autonomous processing are being studied. Autonomous image processing to identify and track objects at the worksite will be

incorporated, which will resolve ambiguities with information from a CAD world model.

Sensors. Sensor upgrades could include laser range imaging, proximity sensors and special purpose sensors.

Operator Interface. The operator's interaction with FTS will be enhanced with stereo vision; video image overlays, including a shaded graphics figure of the robot superimposed on a CAD view of the worksite; displays of sensor data; monitoring of the activity stream at all NASREM levels; speech recognition; and simulation of command sequences.

Safety. System monitoring will be increased to be consistent with the increased autonomy. In addition, a separate "watchdog" system will monitor the main system and output warnings or initiate system shutdown. Monitoring for imminent object collision will be included.

### System Impact

A study [4] assessed the impacts of meeting these requirements:

200 Hz Around the Loop Control Rate. This rate will improve servo control, control loop stability, and force reflection "crispness."

Increased Autonomy. This feature will increase system efficiency and minimize operator load. It will affect processing throughput, worksite modelling, and sensor systems.

Image Data Processing. Processing of image data in real-time requires special purpose hardware and associated software.

Simulation. Requirements for generation of graphic images containing a million polygons at a 6 Hz update rate require extremely fast computers and/or special purpose graphic engines.

CAD Data Interface. This interface with Space Station Freedom will provide the worksite geometric information to support the FTS world model.

## **CURRENT RESEARCH PROGRAMS**

Much of the technology planned for use on the space station FTS and on the test flights is relatively new and untested. To provide the answers needed to design safe, reliable, and fully functional robotics for flight, Goddard Space Flight Center is developing a unique space robotics facility. This highly sophisticated facility will be used by the Goddard robotics team to create, test, and evaluate new robotic technologies required to support Space Station Freedom, and to perform complex maintenance and construction operations for the challenging missions being planned by NASA in the 21st Century.

With help from universities, industry, and other NASA centers, Goddard is presently integrating techniques for task planning, robot motion control, operator interface, sensor integration, advanced end effectors, and safety.

During the past several years, a number of tests and demonstrations were successfully conducted. Some highlights include the autonomous removal of an ORU, assembly of an attached payload structure, vision controlled docking that permits the robot to find an object and dock to it for removal, and voice control of cameras that can free an operator's hands for control of the robot.

Work that is presently underway include detailed task performance with mockups of space station hardware, implementation of higher levels of the NASREM architecture, research into advanced collision avoidance schemes, improvements in operator interfaces, control algorithm development, and prototype work with advanced mechanisms for end effectors, tools and worksite attachment mechanisms.

## **FTS TECHNOLOGY TRANSFER**

As dictated by Congress, a key objective of the FTS project is to enhance the United State's expertise in automation and robotics and, thereby, contribute to this nation's economic competitive advantage. To capitalize on the emerging FTS technologies that are relevant to ground and space applications, Martin Marietta Space Systems Company and the Goddard Space Flight Center have established a Commercial Applications program to disseminate these technologies to industry. This program involves outreach to companies and organization with a potential interest in FTS technologies, including industries, small business, entrepreneurs, the Centers for the Commercial Development of Space, NASA's Industrial Applications Centers, various U.S. government organization, universities, and NASA-sponsored centers. Also, information is being collected in databases which are available to the user community.

Since first-hand interactions between the users and developers of technology are crucial to smooth and timely transfers, a high level of interactive support is being provided to complete the technology transfer process. These processes are being tailored to fit the requirements of the-recipient.

An industrial briefing of the FTS technologies is planned in early December 1990. Interested parties that could benefit from these technologies or any aspect of the commercial applications activity should plan to attend.

## **CONCLUSION**

The FTS provides a versatile capability for long-term space station maintenance. The basic FTS design is now established and plans for the first development test flight in 1992 are well underway. Although the initial design primarily involves the integration of existing technologies, plans for evolution and growth will incorporate the products of on-going research -- bridging the gap between the NASA's research arm and its flight programs. Through the commercial applications program the mechanisms are being established now to transfer the FTS technology products directly into U.S. industry.

## **REFERENCES**

1. McCain, Harry G.: NASA's First Dexterous Space Robot. Aerospace America, February, 1990, pp. 12-30.
2. Lumia, Ronald: Short Term Evolution for the Flight Telerobotic Servicer, GSFC Internal Report, 13 October 1989.
3. Ring, Michael: FTS Requirements for Data Management and Processing, Phase I Report. Advanced Technology and Research Corporation, prepared for GSFC, 23 March 1990.
4. Data Management and Processing Subsystem Alternate Architecture Study, FTS Contract NAS5-30689, Exhibit 2, Task 3, prepared by Martin Marietta Astronautics Group, 15 June 1990.



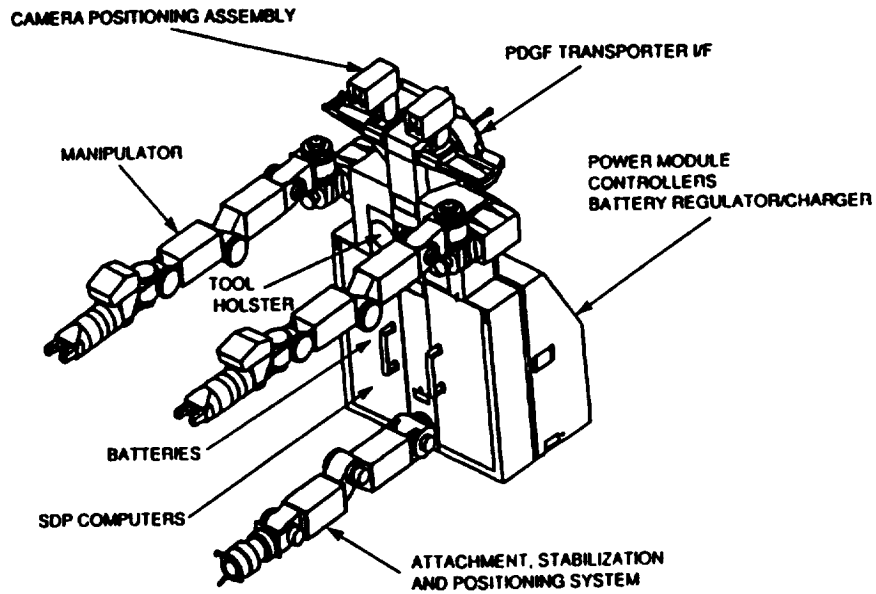
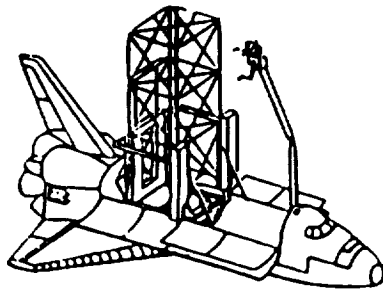
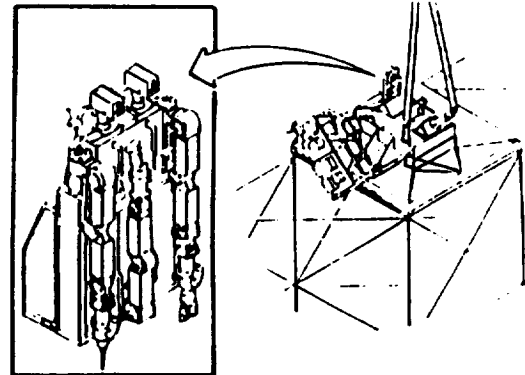


FIGURE 1. THE FLIGHT TELEROBOTIC SERVICER

SSFTS Operations from the Orbiter Payload Bay



Transporter Attached Operations on Freedom



Fixed Base Dependent Operations on Freedom

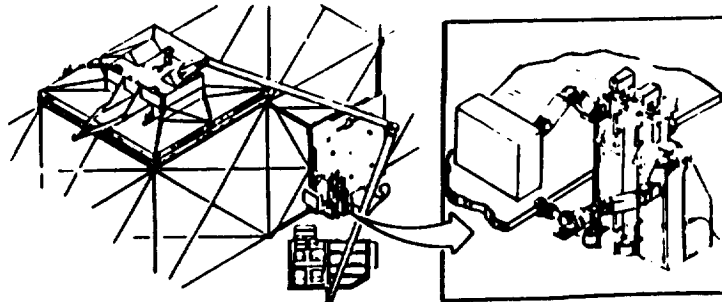


FIGURE 2. FTS OPERATIONS MODE

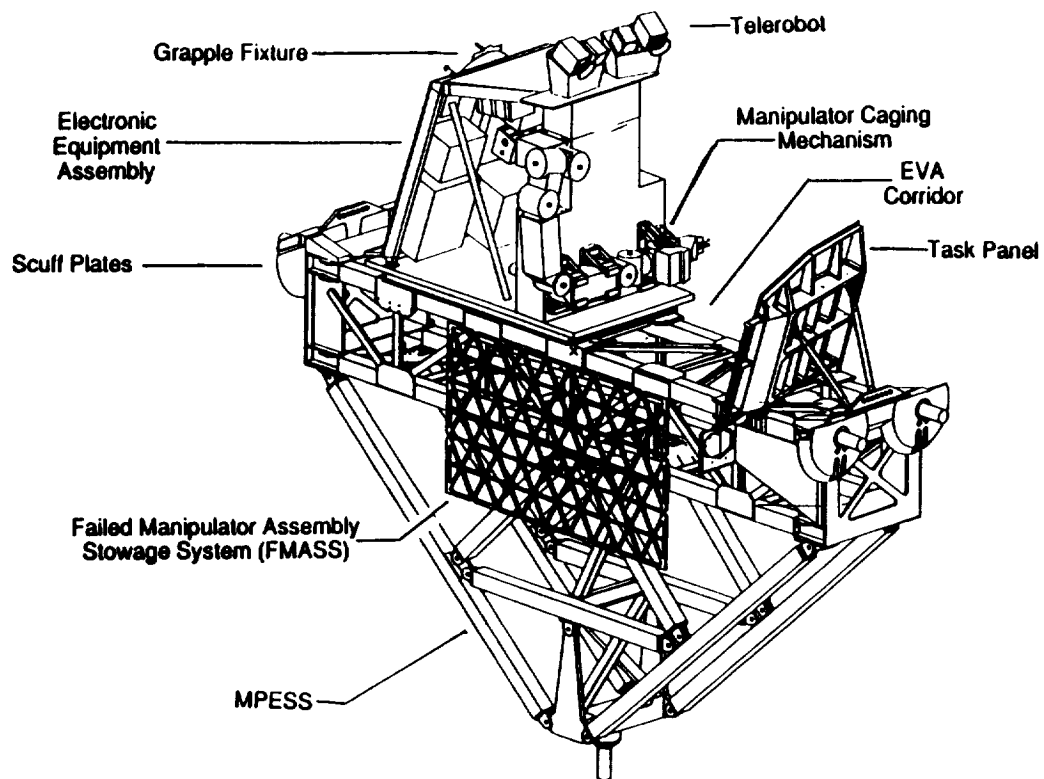


FIGURE 3. DTF-1 CONFIGURATION

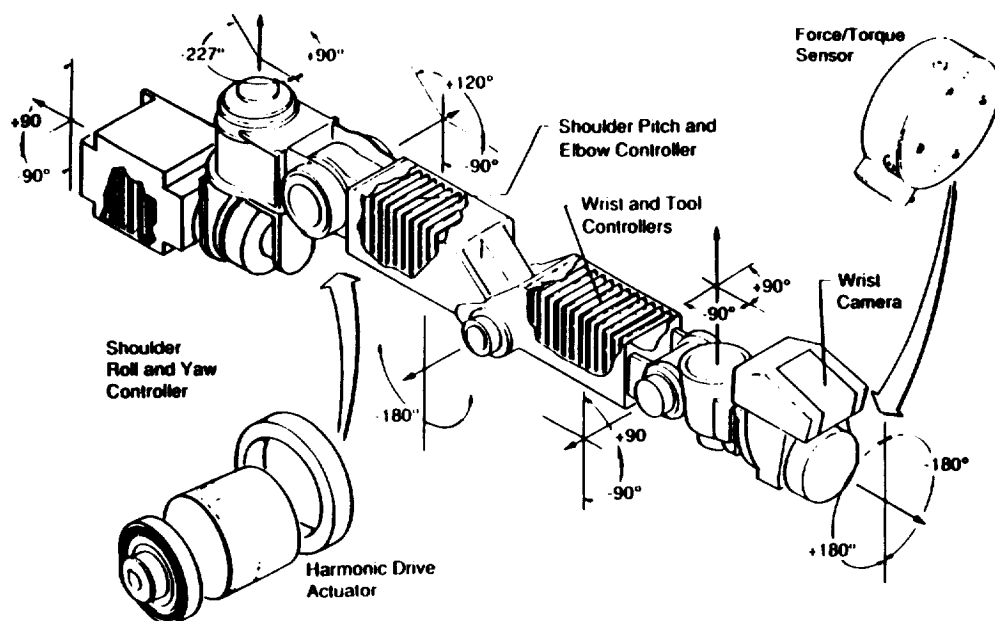


FIGURE 4. FTS MANIPULATOR

- Tip Force At Tool Plate: 20 lb Minimum In Any Direction Or Configuration
- Tip Torque At Tool Plate: 20 ft-lbs Minimum
- Incremental Motion At Tool Plate: <0.001 in, < 0.01deg
- Repeatability At Tool Plate (Constant Thermal Environment): <0.005 in, <0.05 deg
- Translation Rate At Tool Plate, Fully Extended: 24 in/sec Minimum No Load  
6 in/sec Minimum With 2.8 Slug Load
- Accuracy At Tool Plate, Relative To Manipulator/Body Interface: <1.0 in, <3.0 deg
- Actuator Position Sensor Resolution: Shoulder Joints: 22 bits  
Elbow Joints: 21 bits  
Wrist Joints: 20 bits
- All Cable Is Internally Routed
- Manually Releasable Fail-Safe Brakes And Backdrivable Actuators Allow Manual Stowage
- Hardwire Control Allows Backup Operation Of Actuators Using Secondary Motor Windings

TABLE 1. KEY MANIPULATOR CAPABILITIES

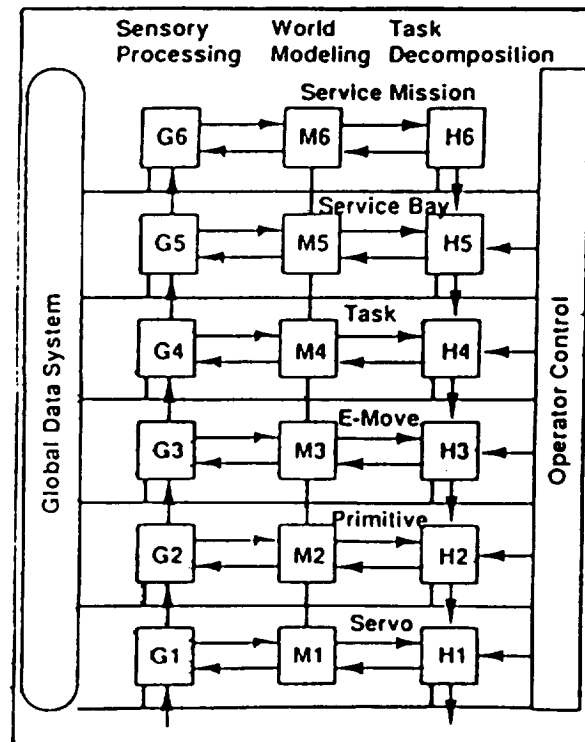


FIGURE 5. NASREM ARCHITECTURE SCHEMATIC

N91-23064

FARMS: THE FLEXIBLE AGRICULTURAL ROBOTICS MANIPULATOR

Paul S. Gill  
Process Engineering Division  
Materials & Process Laboratory  
Marshall Space Flight Center, Alabama 35812

ABSTRACT

A technology utilization project was established with the Marshall Space Flight Center and the University of Georgia (UGA) to develop an earth based, robotic end effector to process live plant (geranium) material which will (A) improve productivity and efficiency in agricultural systems, such as, commercial nursery and green house systems; and (B) apply this technology to NASA's presence in space, including permanently manned space stations and, manned planetary communities requiring large scale food production needs (e.g. vegetable, fruits).

INTRODUCTION

A force-sensing robotic gripper system has been developed at the Productivity Enhancement Complex at the Marshall Space Flight Center. The reason for the robotic end effector development was to alleviate the labor intensive phases of plant production in commercial greenhouse operations.

Due to an increasing competitive world marketplace, automation is beginning to play a key role in the U.S. agriculture industry. The highly advanced growers of Western Europe and the low-cost growers of South and Central America have taken a large portion of the U.S. market in the last several years. For example, the number of cut flower growers in the U.S. has decreased approximately 75% since the mid-1970's in response to imports. In order for the U.S. greenhouse industry to maintain an economically sound industry, new production techniques must be adopted.

Robotics has great potential to meet the need for enhancing the productivity and quality of the U.S. greenhouse industry. Therefore, a robotic workcell has been developed at the University of Georgia Experiment Station, which is also where the MSFC gripper system is being tested and evaluated. The system is being evaluated on its ability to process geranium cuttings. The tedious labor required to process geranium cuttings and the high volume demand for the plant justified an attempt to automate this sector of the greenhouse industry.

The development of this robotic end effector technology was a joint effort between NASA/MSFC and the University of Georgia. The End effector was designed and developed at the MSFC/Productivity Enhancement Complex along with the microprocessor and software to perform the following functions:

1. Determine the cutting (geranium) orientation in perpendicular to the surface (Z-axis)
2. Picking up the cutting using a predetermined pressure
3. Measuring the diameter of the stem
4. Measuring the mass of the cutting
5. Trimming unwanted leaves from the plant
6. Trimming the plant to length
7. Pass cutting thru bend measurement sensor and orient the plant for potting
8. Inserting the plant in a potting trays

The system has been integrated with the UGA agricultural robotics research cell to process geranium cutting for transplanting. The UGA robotic workcell is composed of a 32 bit minicomputer with vision acquisition, an ASEA industrial robot, a parts conveyer and custom fabricated tools. The tasks of the robotic workcell are:

1. To detect cutting in robotic workspace and momentarily stop conveyer.
2. Perform automatic visual analysis.
3. Wait for robot to complete processing previous cutting.
4. Transmit data on cut and grasp locations, stripping requirements, and grade.
5. Automatically signal robot to begin processing, and restart conveyer.

### OPERATION

The gripper system is composed of three subsystems: a closed loop force controlled DC servo actuated two fingered gripper, an industrial PC AT communications node and integral fiber optic analog output sensors used for part feature recognition.

During operation of the cell, geranium cuttings are placed on the conveyor manually. The conveyor indexes each cutting into the field of view of a CCD camera. The camera photographs the part and transmits the vision data to a 32 bit cell computer (Heurikon). The Heurikon calculates the position of the cutting in the XY plane defined by the surface of the conveyor belt. This data is formatted into ASEA robot commands and transmitted to the robot. The robot then moves to a location above the cutting. The arm searches along the negative Z-axis until the axial center line of the part is sensed using the fiber optic distance gauges. The analog output of the fiber optic sensors is sampled by the A/D card residing in the gripper node control. Upon recognition of the part, the gripper is closed using force feedback from a load cell internal to the gripper. Once the programmed force setpoint is obtained by the gripper, the resulting finger opening is stored by the gripper node control. This system caliper data is available to the Heurikon when needed.

The part is grasped by the gripper and carried by the arm to a trimming station. The trimming station removes the two petioles at the base of the stem and trims the base of the stem; the part is then placed in a fixture on an electronic balance. The balance is polled for the mass of the part by the gripper node control. This mass data is also available to the Heurikon when needed. The gripper is closed under force control a second time. The part is grasped by the gripper and carried by the robot arm to a stem alignment measurement station. Data from this station is used by Heurikon to command the robot to correctively place the stem into a potting tray. Upon completion of the cycle, the gripper is homed. The next cutting is positioned under the camera by the conveyor and the process is repeated.

### GRIPPER/CONTROLLER DESIGN

#### HARDWARE

The force-sensing robotic gripper system contains four basic modules, namely the gripper, the Telerobotics controller enclosure, the Delta Tau SMCC-PC motion controller card, and two Banner photoelectric sensors (attached diagram).

The gripper consists of a precision servo motor driven, two finger, linkage parallel unit. The gripper uses an encoder for position indication and a miniature precision load cell for sensing the closure force between the two fingers. The gripper has a force range of 0 to 30 lbs with a force resolution of +/- 0.5 lb. The maximum closure rate is approximately 6 inches/second. The finger opening width is 4 inches. The gripper weighs approximately 6.5 lbs.

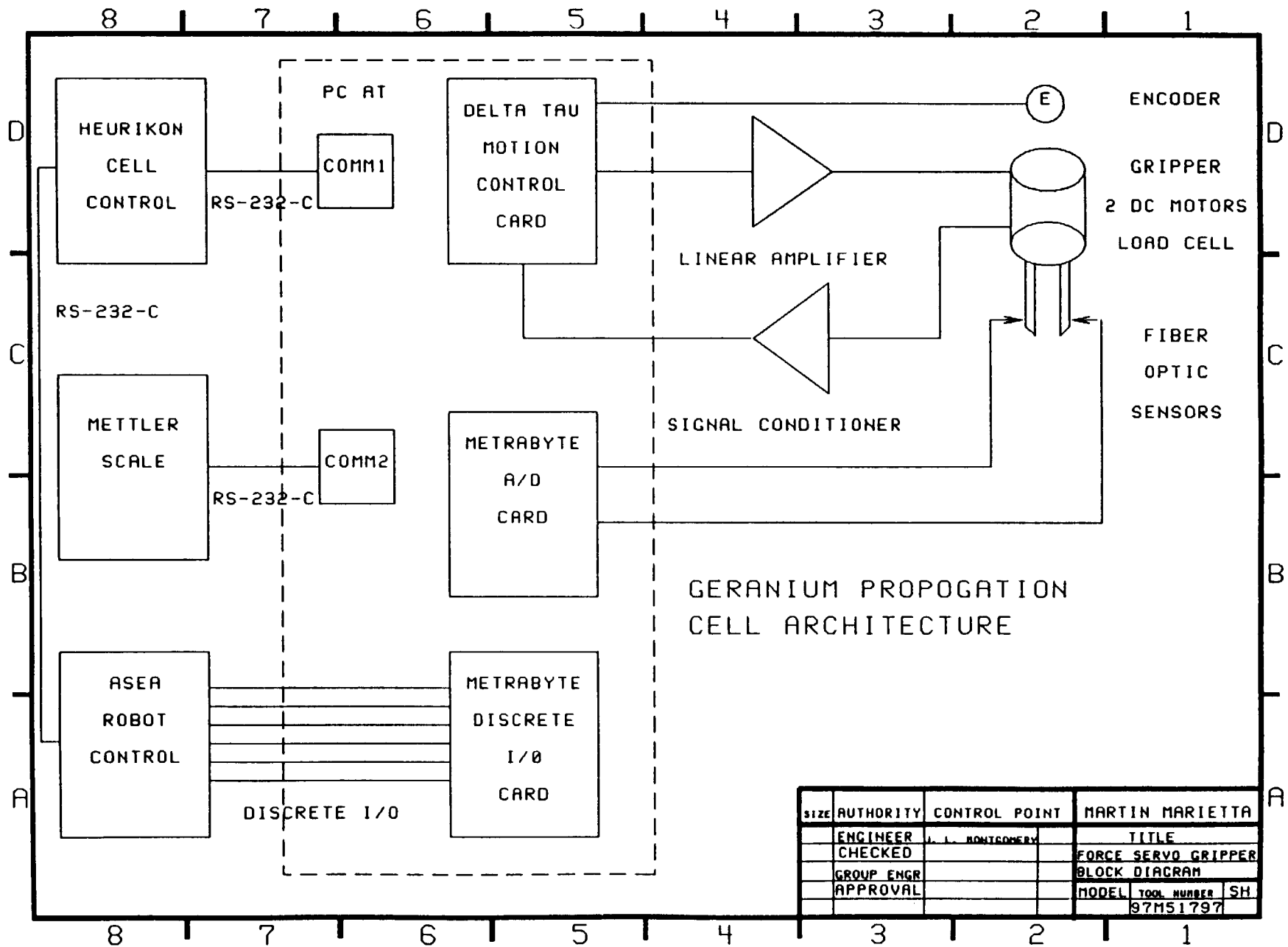
The gripper controller contains a driver-amplifier, a motor power supply, a low voltage power supply, and a signal conditioning board. The signal conditioning board prepares the load cell's signals for the motion control card contained in the PC. The motion control card is a full two axis digital PC card capable of receiving a variety of position, velocity, and analog signals. These signals are digitally processed by the system's software. The controller generates motor commands which in turn are executed and amplified by the driver-amplifier.

Due to the nature of the geranium cutting, the leaves of the plant forces the stem to be positioned slightly above the conveyor. Therefore, photoelectric sensors are used to assure that the robotic fingers will close directly onto the stem. The photoelectric sensors used are analog infrared scanners that provide a variable DC voltage output that is proportional to the strength of the received signal. The reason that the sensors are analog is that the gripper has to have the ability to distinguish between the stem and a branch of the geranium. If the sensor was digital and a branch interrupted the infrared beam, then the robot would stop prematurely. The sensors located on the finger tips of the gripper must be able to position themselves between the branches and stem of the plant. Therefore, the cables for these sensors have right angle deflectors.

## SOFTWARE

The end effector controller system is a comprehensive group of IBM AT compatible based program modules developed and written in "C" language programming environment and a series of programs written in a language specific to the Delta Tau Motor Controller Card. The "C" program modules are combined in a task designed to initialize the hardware interfaces, provide a comprehensive variety of displays for operator interaction, and process a series of interactive commands received from a host processor or a robot controller to manipulate an end effector. The motor controller programs are kept in files on disk and are loaded into the motor controller card to perform the gripper functions.

Upon entry into the system, the interfaces, serial port (COM1) for communications with the host processor and serial port (COM2) for communications with the scale, are initialized, the optics interface, an A/D converter (DAS-16), the end effector interface, a motor controller (SMCC), and the robot interface, a parallel discrete I/O (PIO-24), are also initialized. The screen then displays the status window indicating the initial main menu node. From this point on, the system is very user friendly and in each of the selections provides the operator with a series of submode menus to guide the operator through the process in an orderly fashion and with descriptive information and data displays for operator viewing. The remote operations selection from the main menu is the function that performs as an end effector controller and processes the commands received from the robot controller or the host processor. The commands to manipulate the end effector are performed by downloading a program into the motor controller card and running the program on the card. Command completion and information requests are returned to the initiator as required. The remaining selections from the main menu deal with test modes for the individual peripheral interfaces and the changing of parameters utilized by the end effector during remote operation.



## CONCLUSION

The design, fabrication, development, testing and evaluation of the force-sensing robotic gripper system was performed by NASA and Martin Marietta personnel at the Marshall Space Flight Center. The implementation of the gripper system into the University of Georgia's robotic workcell was successful. Test results indicate that the force-sensing gripper does not crush or bruise the plant. The results also show that the percent of successful robot cycles is approximately 90 percent. However, the photoelectric sensors slow the robot's cycle time by 1 to 2 seconds which gives the workcell a cycle time of 6 to 7 seconds.

The development of the robotic workcell is still in process at the University of Georgia. The MSFC is supporting any enhancements or modifications to the gripper system and the robotic workcell. Modifications to the workcell's vision system and feeder system is currently in process. As soon as the workcell has been optimized in the laboratory, it will then be installed in the greenhouse for further testing. If the results from these tests are successful, then this technology may be applied to greenhouses across the U.S. or even in a greenhouse in a prolonged space habitat.



### ACKNOWLEDGEMENTS

This project was initiated by Dr. Ward Simonton of the University of Georgia (UGA) who, together with his associates helped support the design and development of the Flexible Robotic Manipulator System and integrated the manipulator into the robotic workcell of (UGA) Experiment Station, Ga.

Jim Montgomery, John Shiver and Neil Heffelfinger and associates of Martin Marietta supported our in house effort in the design and development of the Flexible Agricultural Robotics Manipulator System.

## DIVERSE APPLICATIONS OF ADVANCED MAN-TELEROBOT INTERFACES

Douglas A. McAfee

Robotic Hardware & Flight Experiments Group  
Jet Propulsion Laboratory  
California Institute of Technology  
4800 Oak Grove Drive, 138-212  
Pasadena, CA 91109

### ABSTRACT

Advancements in man-machine interfaces and control technologies used in space telerobotics and teleoperators have potential application wherever human operators need to manipulate multi-dimensional spatial relationships. Bilateral six degree-of-freedom (6-DOF) position and force cues exchanged between the user and a complex system can broaden and improve the effectiveness and intuitiveness of several diverse man-machine interfaces.

### A LOOK AT JPL MAN/TELEROBOT INTERFACES

JPL has developed several man-machine interfaces that capitalize on the primary human communication channels, visual, tactile, and auditory, to transfer information to and from a telerobot.

Much of the information fed back to the operator is visual. An operator's control station, like those shown in Figures 1 and 2, typically provides multiple TV views of the remote worksite and may include a stereoscopic view as well for better depth perception.

A computer generated graphics display provides the operator with force/torque vectors representing the contact interactions between the telerobot manipulator and it's environment. There are also various other menu driven graphical interfaces that allow the system programmer, or even the operator, to easily adjust a broad spectrum of system parameters, such as gains, pot and encoder biases, etc., to accommodate personal preferences or system stability requirements. There may also be advanced multi-window graphical displays for implementing and pre-checking high level AI/Task Planning activities [1,2,3].

There may be situations where there are excessive time delays between the commanded input to telerobot and it's actual occurrence. For example when the operator and the remote manipulator is separated by exceptionally long distances or when there are excessive computational complexities. In these circumstances the operator control station may be equipped with high speed predictive displays where graphical overlays of "phantom robots" are superimposed over the actual video images returned from the remote worksite. These phantom graphic images are updated in real time and show the impending motions ahead of time and can help reduce confusion [4].

There are also voice recognition systems, standard computer terminals, key boards, mice and joysticks, and numerous lights, buttons, switches, alarms, buzzers, and beeps to keep track of.

In the midst of all this, the operator is furnished with a unique and highly intuitive tool that provides him immediate and direct control of the remote telerobot manipulator. This device is known as a force-reflecting hand controller and will be the primary focus of this paper along with several of it's potential spin-off applications.

### JPL HAND CONTROLLERS

By simply grasping the handgrip of the hand controller and moving it in the desired direction, the operator has direct control of the remote manipulator's position, and/or speed and direction. If the arm

comes in contact with some object at the remote worksite the resulting forces and torques are transmitted back through control system and the hand controller mechanism to the human operator's hand thereby allowing him to sense the remotely applied forces. The complex transformations and servo control loops required to perform such an operation are all performed completely transparent to the human operator. He can perform the remote task almost as if he were performing it in person. Many experiments have shown that when a human operator is provided with force feedback information the task completion times and extraneous contact forces are greatly reduced [5].

Beginning in the late 1970's, JPL developed a series of FRHC's which have been installed in several research projects over the years including JPL's Telerobot Testbed Facility and at the Advanced Teleoperator Research Lab (shown in Figures 1 and 2 respectively). These early JPL hand controllers, dubbed the Models A, B, and C, have collectively undergone many hundreds of hours of testing and operation and have been the means of greatly expanding the scientific and engineering knowledge about man-machine interfaces as they apply to bilateral teleoperation of robotic manipulators. Extensive and sophisticated control methodologies have been developed based on these hand controllers [6,7,8].

In 1988 initial funding was provided to develop a more robust flight-qualifiable force-reflecting hand controller (FRHC). In addition to meeting the rigorous demands of space flight it was also to have a smaller stowage volume and much higher mechanical bandwidth than its predecessors. After several months of intense design and development a pre-flight prototype FRHC, called the Model X, was produced (see Figures 3 and 4). The basic design, physical description, capabilities and the results of early testing are fully described in [9].

#### **BASIC FEATURES OF THE MODEL X FLIGHT HAND CONTROLLER**

The Model X FRHC has many outstanding features which combine to make it a substantial step forward in hand control design. Several of the features are unique to the Model X and promise to improve the performance bandwidth of the human operator in the control of remote applications.

At first glance some features may not be immediately obvious or impressive to someone unfamiliar with the subtleties of hand controller design. Where possible, the importance of these, less obvious, features have been highlighted. Some of the features were measured experimentally while others values were derived analytically or approximated.

- 1) **Highly Intuitive Operation.**  
The remote tasks are performed as if in the operator's own body-reference frame. Teleoperation complexities are largely transparent to the operator. The design inherits sophisticated control modes from predecessors.
- 2) **Generalized Applicability.**  
The Model X can be use to command diverse remote manipulator's that are kinematically dissimilar to itself.
- 3) **Excellent 6-DOF Position and Orientation Resolution.**  
With the Model X fully out stretched, in it's worst-case position, it can detect and respond to positional inputs at the handgrip as small as 0.002 of an inch.
- 4) **Excellent Dynamic Force Range.**  
In its present implementation, the prototype Model X can output feedback forces to the human operator ranging from = 3 to 67 oz. (0.2 to 4.2 lbs) when fully extended in 1-g (worst-case). When less than fully extended it can output forces as high as 12 lbs. Also, the force output levels in the zero-g of space would be substantially higher.

- 5) **High Mechanical Stiffness and Fast Mechanical Response.**  
The Model X is a much stiffer mechanism than other current hand controller designs with an approximate natural frequency of around = 25 hz. With this higher mechanical bandwidth the Model X can reproduce higher fidelity force feedback cues to the human operator, making the feedback more crisp and distinguishable.
- 6) **Flight-Qualifiable.**  
The Model X has a rugged structural design which is necessary for the rigorous demands of space flight.
- 7) **Low Mass.**  
The prototype Model X mechanism has a mass of only 27.6 lbs. And further weight reduction is possible.
- 8) **Small Stowage Volume.**  
When not in use the Model X can fold up and stow in a relatively compact case measuring 9.6 x 17 x 20 inches (1.9 ft<sup>3</sup>)
- 9) **Large Workspace.**  
At over 4 ft<sup>3</sup> the Model X has a much larger work envelope than previous FRHC designs. Reduced indexing required.
- 10) **Range of Motion Optimized to Human.**  
The Model X's work envelope overlaps much of the natural reach of the person operating the hand controller. The mechanical layout is such that it shadows the movements of the human operator and does not interfere with the viewing and operation of other control station equipment.
- 11) **Low Friction.**  
It has relatively low friction levels ranging from 3.0 oz. to 6.0 oz. within target areas of the work envelope.
- 12) **No Backlash.**  
The Model X design uses a unique cable/pulley transmission system that eliminates all backlash. No gears are used.
- 13) **Low Inertias.**  
The Model X has a unique design that permits all six motors, constituting the greatest portion of the mechanism's mass, to be mounted at or near the hand controller's base thus greatly reducing the effects of inertia that may otherwise be detrimental to the Hand Controller's useful operation.
- 14) **Highly Backdrivable.**
- 15) **No External Counter Balancing Required.**
- 16) **Fully Decouple Joint Motion.**  
Each of the six joints on the Model X rotates independently from all others eliminating additional computations otherwise required. It also reduces the complexity of dynamic and friction compensation techniques.
- 17) **Simple and Computationally Fast Kinematic Solutions.**  
The individual link coordinate reference frames either intersect at right angles or are parallel to one another. This makes many of the terms in the coordinate transformations reduce to

either 1 or 0 and thereby greatly simplify their computations. It also has no link offsets or link twist angles to factor in.

The foregoing list of features provides many indications that the Model X hand controller will be a viable and effective user interface to spatially dynamic multi-degree-of-freedom systems.

## **SIX POTENTIAL SPIN-OFF APPLICATIONS FOR THE MODEL X HAND CONTROLLER**

The Model X hand controller is a highly intuitive, highly versatile human interface to dynamic multi-degree-of-freedom machines. As a high resolution 6-DOF position input device with exceptional force feedback capabilities it may prove to be a useful human interface in applications other than the control of teleoperated robotic arms. The Model X has been optimized to fit the human operator's range of motion. As such, it has the potential of being the fundamental interface that a human operator needs in order to manipulate multi-dimensional spatial relationships where force cues can help associate a coordinated response.

### **SINGLE-HAND FLIGHT CONTROLLER FOR HELICOPTERS**

Helicopters are undoubtedly one of the most difficult and challenging vehicles to pilot, especially during the high agility maneuvers required in modern combat. Many research efforts have been undertaken to reduce and optimize the workload demanded of military helicopter pilots [10, 11, 12].

One possible spin-off application of the Model X force-reflecting hand controller technology, may be in the field of helicopter pilot flight control interfaces. Advances in digital flight control techniques and fly-by-computer technologies are opening the door for sophisticated new pilot/vehicle interfaces that can potentially significantly simplify flight path management and thereby allow the pilot to concentrate more on the overall mission management objectives.

As a flight control interface, the Model X hand controller would allow the pilot to command even the most complex helicopter flight path using only one hand, whereas the four conventional pilot inputs generally in use today require the use of both hands and both feet (see artists concept Figure 5).

The Advantages include: a) Significant Reduction in workload through highly intuitive single-hand flight control; b) decreased cockpit congestion; c) reduced training time; d) improved maneuverability and handling qualities; e) highly adjustable performance attributes; f) easy system upgrades and modifications; g) accommodates individual pilot preferences and modular mounting configurations; h) seamless auto/manual mode transitions; i) decreased vulnerability and complete redundancy possible; j) lighter weight; and, k) less expensive than conventional controls.

### **REMOTELY CONTROLLED UNDERWATER ROBOTS**

Without a doubt, the wave of the future in ocean engineering and science are remotely controlled unmanned vehicles. As a highly intuitive and compact man/machine interface, the Model X hand controller has multiple potential applications within the field of remotely operated vehicles (ROV's).

In Recent years we have seen the emergence and evolution of several new underwater research vehicles like Woods Hole's "Jason", Eastport's "Gemini 11", Hydrovision's "Hyball", and NOSC's "Advanced Tethered Vehicle". Some of the more advanced remotely operated vehicles (ROV's) can be maneuvered in 6 DOF (i.e., forward/back, up/down, side-to-side, roll, pitch, and yaw) and include sophisticated navigation modes such as auto-heading, auto-depth, and position hold.

The ROV's often carry an array of sensors and tooling such as: still and video cameras, lights, scanning sonar, rate gyro's, temperature sensors, salinity sensors, etc. In addition, some unmanned submersibles may also employ one or more robotic manipulators and grippers (see Figure 6).

There are three areas in which the Model X force-reflecting hand controller might conceivably be used to enhance the man/machine interfaces of advanced underwater exploration robots. These areas are: (1) remote piloting of the unmanned, underwater robotic vehicle; (2) remote control of the robotic manipulators mounted on the vehicle; and, (3) remote control of camera pan, tilt, zoom, and focus. All integrated into a single hand controller.

### **NUCLEAR INDUSTRY REMOTE MATERIALS HANDLING**

Since the 1940's the nuclear industry has developed and used a wide variety of hand controllers. Many of the recent advancements in teleoperation and telerobotics can be directly attributed to research efforts in this field. The nuclear industry, both private and military, employs most of the hand controllers in used today which are typically mechanically coupled master/slave type devices.

The nuclear industry has many applications where teleoperation is essential both from a safety and an economics standpoint. Clean-up or even routine assembly and maintenance operations are composed of tasks that are far too complex to be performed by unsupervised autonomous robots and too dangerous to be done by human workers. Because of these issues the industry has turned to a legion of remotely operated devices.

A single, generalized hand controller, like the Model X, could be used to control an unlimited variety of remote manipulation devices that differ substantially in both structure and kinematics. Devices like overhead cranes, spanning hundreds of meters, used to move equipment or pallets of radioactive material; or small dexterous robot manipulators used inside hot cells to assemble fuel canisters. The operator control stations for these various tasks could all utilize a common type of man- machine interface.

### **FLIGHT CONTROL OF REMOTELY PILOTED VEHICLES (RPV's)**

RPV's are typically envisioned as high-tech sentinels, pinpoint target acquisition instruments, aircraft/missile decays, and covert surveillance platforms [13].

Most experimental RPV's are designed to fly like an airplane but some are capable of more complex flight trajectories more like a helicopter. They can hover, fly forward/back, up/down, side-to- side, and perform roll, yaw, and pitch attitude control maneuvers. Presently RPV flight paths are typically controlled by either pre-set programs or joysticks similar to radio controlled model planes.

The Model X could be employed as the man-machine interface through which a human operator remotely pilots an RPV. The mechanics of this new type of pilot/vehicle interface would essentially be the same as that described for helicopter flight controls except that here the vehicle's pilot would not be physically on the vehicle being flown. It could easily be made into a light weight and backpackable.

### **SIX-DOF COMPUTER IMAGE MANIPULATION WITH FORCE FEEDBACK CUES**

Many advanced computer graphics packages allow the user to create three dimensional objects. These objects can then be manipulated relative to the user's point of view and to each other.

Today CAD programs often allow the user to zoom in, zoom out, and move graphical objects up, down, or sideways relative to the screen. A user can also rotate an object about an arbitrary axis. There have been several new devices that attempt to make this object manipulation more user friendly and less cumbersome, like 3-DOF joysticks, "advanced mice" and track balls [14], but they often limit manipulation to only one, two, or three dimensions at a time instead of the six required to pinpoint any object's position and orientation (i.e., X, Y, Z, Roll, Pitch, Yaw).

Using the Model X, computer users would be able to move and re-orient computer graphic images as easily and as intuitively as if they had reached out and done so with their own hand. It would be like having the functionality of six joysticks rolled into one.

Add to this the capability of the Model X to output feedback forces and one has a completely new level of computer interface. The user would then be able to "feel" objects as well. This would allow a wide range of new uses. For example, force cues could be given to the user when objects being manipulated on the screen contact one another, or there could be virtual buttons and switches having actual mechanical detentes, or the user might even sense the relative mass and inertias of objects. The capability for the computer to output forces could greatly expand the sensory bandwidth of the human interface at graphics workstations, creating a virtual environment with the added dimension of dynamic force.

A more specific example might include research into new chemical compounds is being performed on computers. It is sometimes very beneficial to the researcher to be able to visualize complex chemical chains in 3-D and then to manipulate and mate new chains at various locations by hand. As explained in the previous application the Model X has the potential of greatly enhancing the computer graphic interface. The force feedback capability might be employed to provide the researcher with a sense of feel for the structure of the microscopic substances he is working with and provide additional feedback about the connectivity of various complex substances. Appropriate attractive and repulsive forces could be applied to particular molecules.

One researcher that studies the complexities and spatial relationships of DNA and other genetic materials has sought to augment the laborious process of checking potential DNA combinations and fits by using CAD rendered displays and a hand control device [15]. By adding the dimension of force feedback and the simulated molecular docking process might be enhanced even further.

### REDUNDANT DOF DEXTEROUS ROBOT MANIPULATORS

There is a significant research effort under way within the robotics community to develop new methods for controlling redundant degree-of-freedom (DOF) manipulators, particularly robots with seven or more DOF.

From extensive earlier work with 6-DOF manipulators (which is the minimum number of degrees-of-freedom necessary to provide complete position and orientation capability) we learned that several features inherent in the mechanism such as: kinematic simplicity, friction, manipulator stiffness, mass distribution, backlash, backdrivability, and others, were important issues to consider during the mechanical design of the robot. If they were not dealt with properly during the mechanism design then they could be much more cumbersome and/or impossible to compensate for later on in the software control algorithms.

The technologies used in the design of the Model X hand controller could be used to design new redundant DOF manipulators that had excellent position and force control characteristics.

### SUMMARY

Many engineers, designers and researchers in several fields have understood the need for better man-machine interfaces. Following an informed discussion on the state-of-the-art in teleoperation in a recent Scientific American article, Dr. William Uttal, professor of psychology at Arizona State University, concluded by making the following observation:

"These prototype teleoperators show that most of the physical technology needed for these highly complex and useful teleoperators is already available. **The major challenges now lie at the interface of human and machine.** Until sufficient progress in computer technology makes it possible to substitute artificial intelligence for human intelligence, psychological insights into the best ways to link human operators to machines will pace the evolution of remote workstations" [16].

The Model X FRHC is a significant step forward in hand controller design and may prove to be an extremely useful new man-machine interface in several diverse applications including: (a) new pilot/helicopter flight control interfaces; (b) control of underwater remotely operated vehicles, manipulators, and other system components; (c) remote materials handling, assembly, maintenance, and inspection operations in the nuclear industry; (d) new pilot/RPV flight control interfaces; (e) 6-DOF computer image manipulation with force

feedback cues to the operator; and, (f) design of redundant DOF dexterous robot manipulators with good force controllability, minimal kinematic complexity.

### ACKNOWLEDGMENTS

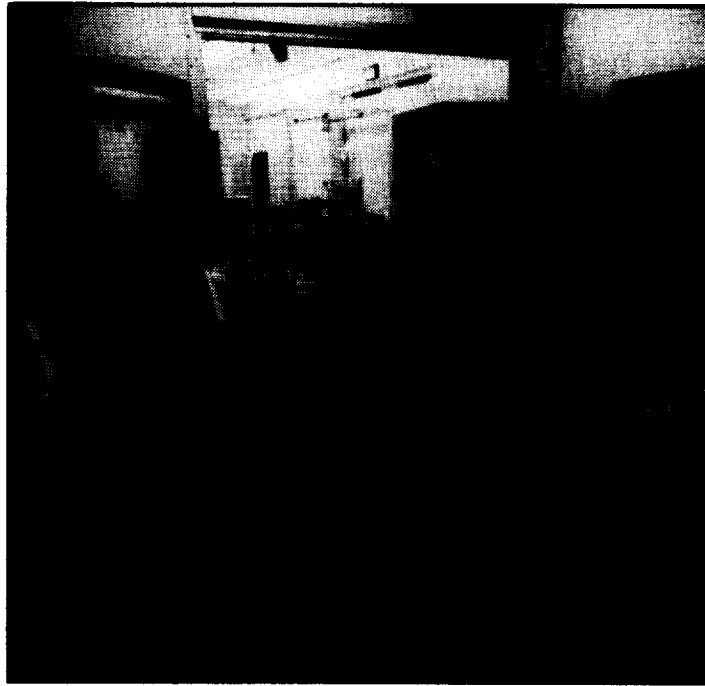
The technologies described in this paper were carried out by the Jet Propulsion Laboratory, California Institute of Technology under contract with the National Aeronautics and Space Administration. The author would also like to recognize the significant contributions of Dr. William T. Townsend and Edward Snow as co-designers of the Model X mechanism and the efforts of Brad Gibson, Brad Swenson, Brian Okerlund and Richard Fleischner from the JPL CAD Services Group.

### REFERENCES

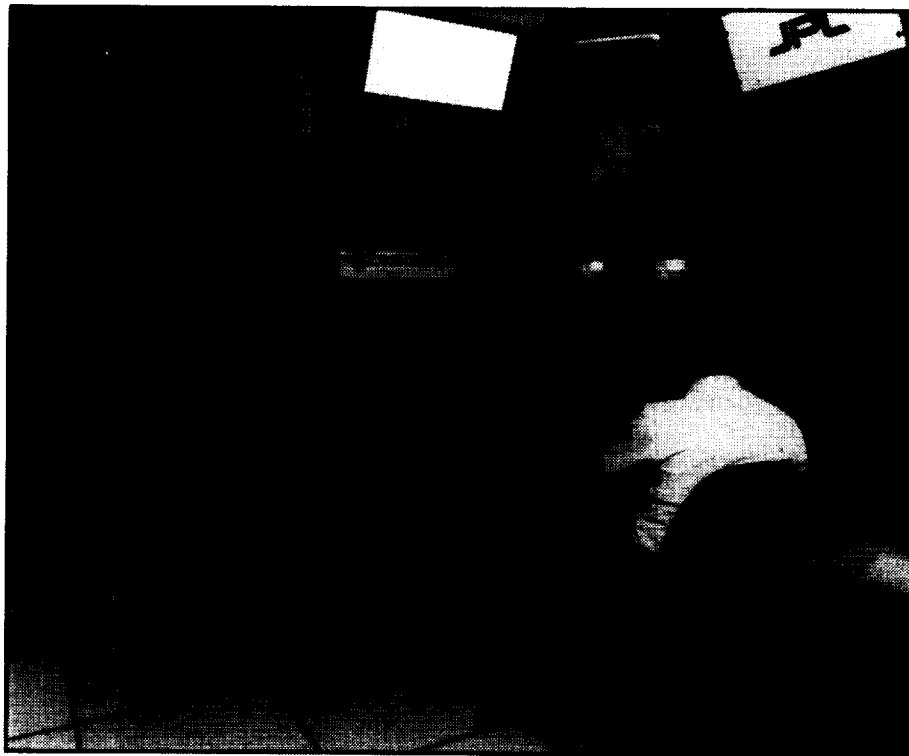
- [1] The JPL Operator Control Station, Ed Kan, et al, Proceeding of the NASA Conference on Space Telerobotics, January 31, 1989, JPL Publication 89-7, Vol 11, pages 51-74.
- [2] A Testbed for a Unified Teleoperated-Autonomous Dual-Arm Robotic System, Samad Hayati, et al.
- [3] UMI: An Interactive Supervisory and Shared Control System for Telerobotics, Paul Backes and Kam S. Tso, Proceeding of the 1990 IEEE International Conference on Robotics and Automation, May 13-18, Cincinnati, OH, pages 1096-1101.
- [4] The Phantom Robot: Predictive Displays for Teleoperation with Time Delays, Antal Bejczy, Won S. Kim, and Steve Venema, Proceeding of the 1990 IEEE International Conference on Robotics and Automation, May 13-18, Cincinnati, OH, pages 546-551.
- [5] Performance Evaluation of a Six-Axis Generalized Force-Reflecting Teleoperator, Blake Hannaford, et al, JPL Publication 89-18, June 15, 1989.
- [6] Controlling Remote Manipulators Through Kinesthetic Coupling, A.K. Bejczy and J.K. Salisbury, Computers in Mechanical Engineering, July 1983.
- [7] Teleoperator Subsystem/Telerobot Demonstrator: Force-Reflecting Hand Controller Equipment Manual, Douglas A. McAfee and Timothy Ohm, JPL Document D-5172, January 1988.
- [8] Performance Capabilities Of A JPL Dual-Arm Advanced Teleoperation System, Z.F. Szakaly and A.K. Bejczy, SOAR '90 Workshop, Albuquerque, NM, June 26-28, 1990.
- [9] Model X Force-Reflecting Hand Controller Mechanical Characterization and Initial Evaluation, Douglas A. McAfee, JPL Document D-7250, March 1990.
- [10] An Advanced Digital Flight Control Concept For Single Pilot, Attack Helicopter Operations, Charles A. Parlier, 42nd Annual Meeting of the American Helicopter Society, 1986.
- [11] A Comparison Between Various Side-arm Controller Configurations in a Fly-By-Wire Helicopter, J.M. Morgan, 44th Annual Forum of the American Helicopter Society, Washington D.C., June 1988.
- [12] Helicopter Flight Control With One Hand, A.L. Lippay, et al, Canadian Aeronautics and Space Journal, Vol 31, No. 4, December 1985.
- [13] Air Force UVS 2000, P. Cassibry, et al, 12th Annual Meeting of the Association for Unmanned Vehicle Systems, Anaheim, CA, July 1985.
- [14] Visualizing In Three Dimensions, Ric Dehoft, Electronic Engineering Times, July 31, 1989.



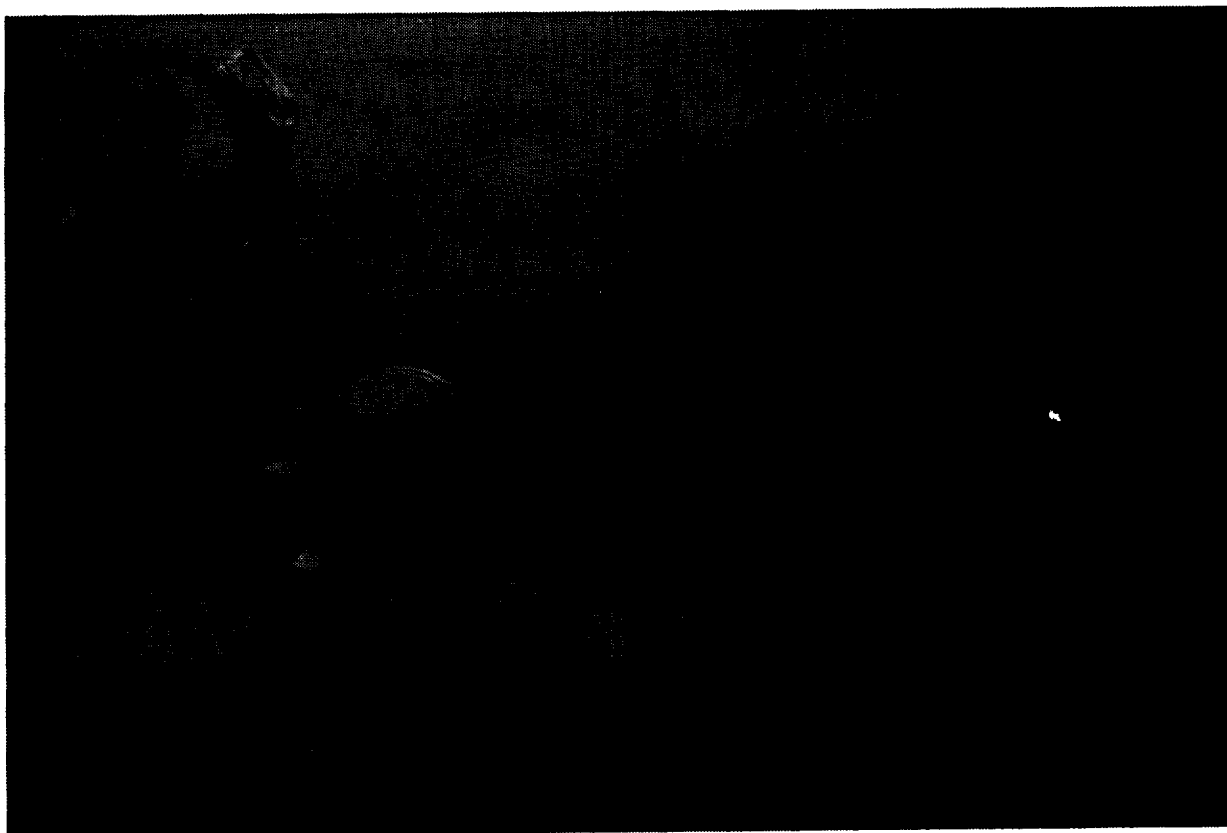
- [15] Using a Manipulator For Force Display in Molecular Docking, Proceedings of the IEEE Conference on Robotics and Automation, April 1988.
- [16] Teleoperators, William R. Uttal, Scientific American, Dec 1989.



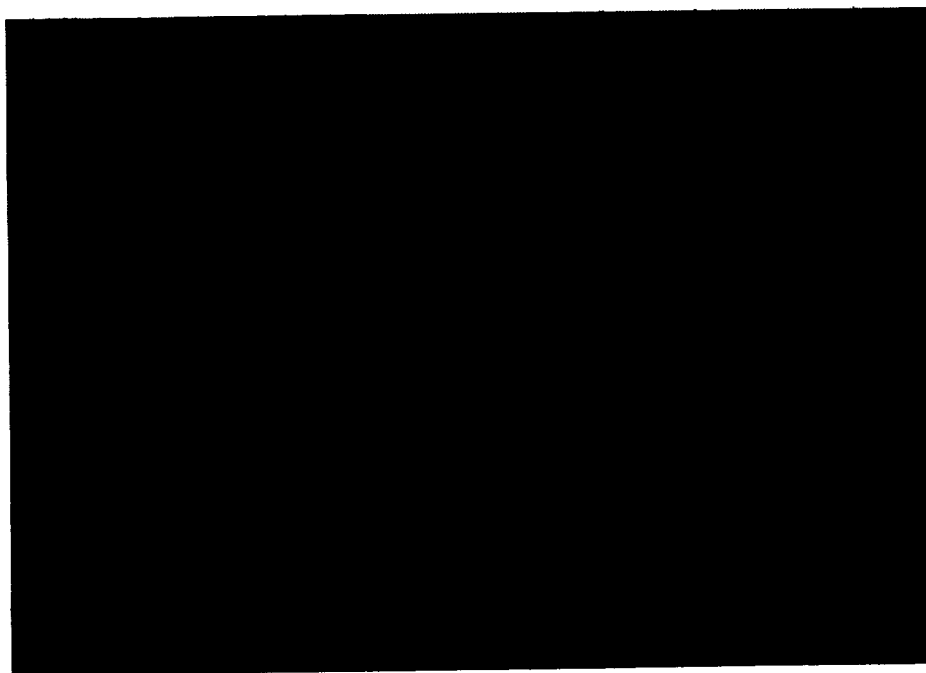
**Figure 1**      **The Advanced Teleoperator Research Laboratory**  
Showing the Model B FRHC, three camera views, a force/torque display, and a predictive display. The remote telerobot and task board are visible in the background.



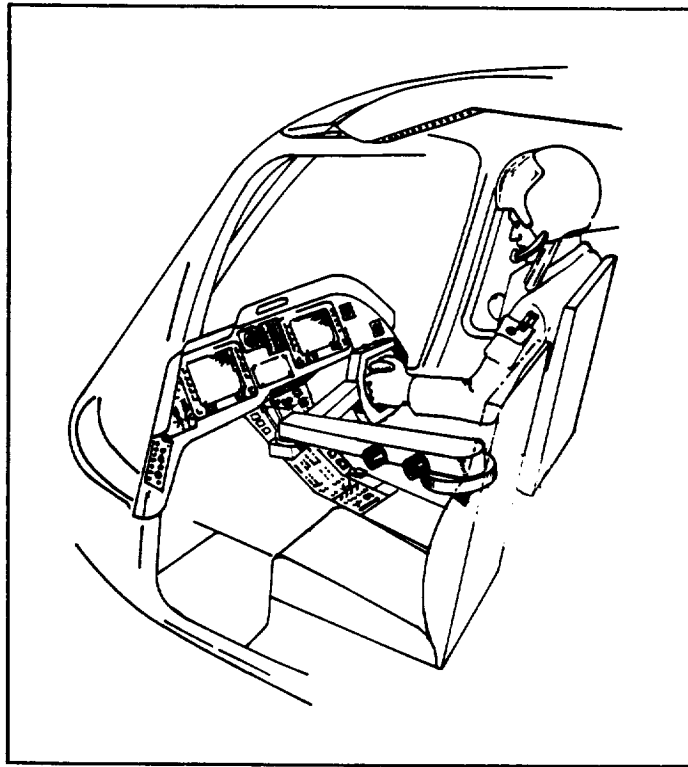
**Figure 2**      **Telerobot Testbed Facility**  
The basic operator control station has five separate camera views, plus a stereoscopic view, two Model C FRHC's, a graphical multi-window user interface, and a voice recognition system.



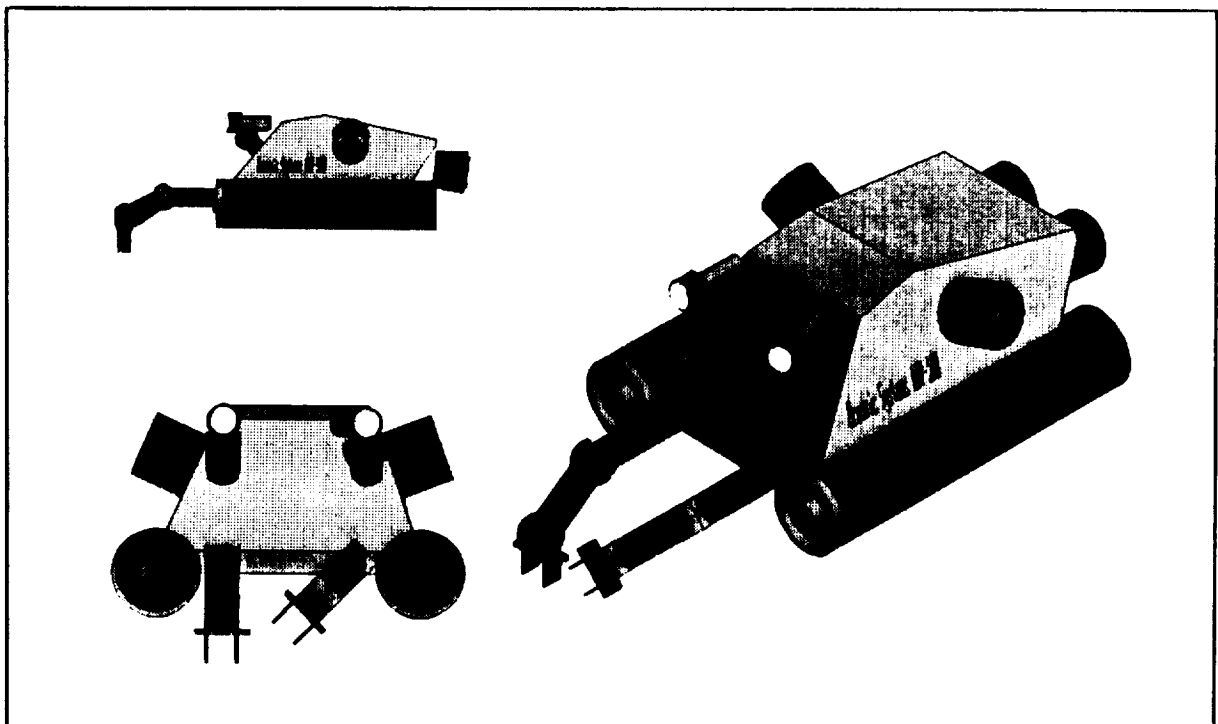
**Figure 3      The Model X Force-Reflecting Hand Controller  
Shown in the vertical mounting configuration.**



**Figure 4      The Model X shown in the horizontal mounting configuration.**



**Figure 5** The Model X Hand Controller could potentially be used as a radically new 6-DOF pilot/helicopter flight control interface supporting advanced digital flight control systems.



**Figure 6** The Model X Hand Controller could be used to pilot this underwater vehicle and control it's robotic arms and cameras.

## **SESSION H - SENSORS AND MEASUREMENT TECHNOLOGY (PART 1)**

**Tuesday November 27, 1990**

- **Urodynamic Pressure Sensor**
- **Electron Tunnel Sensor Technology**
- **Practical Approaches For The Application Of Resistance-Type Strain Gages On High-Temperature Composites**
- **AI Mass Spectrometer For Shuttle Health Monitoring**
- **Instrumentation For Optical Ocean Remote Sensing**
- **Monitoring And Control Of Atmosphere In A Closed Environment**



## URODYNAMIC PRESSURE SENSOR

Thomas Moore  
Aerospace Engineering Technician  
NASA Langley Research Center

### Introduction

A widespread method for measuring the closing pressure profile along the female urethra is to insert a sensor into the bladder and withdraw it at a controlled rate such that time translates into distance and sensor output or pressure is the dependent variable. A second sensor element, located distally from the first, remains in the bladder to provide a reference pressure. By synchronizing the speed of an analog strip chart recorder with the withdrawal rate of the sensor, one obtains a measure of the pressure profile along the urethral length. Two aspects of this method are controversial, first, because the sensor area is quite small, it includes only a relatively small portion of the total circumference giving rise to questions regarding circumferential sensitivity, and, second, as the sensor is being withdrawn, transient events, coughs, muscular impulses, etc., cannot be duplicated over time in order to examine the response of the total organ to a single event. With these concerns in mind, a transducer system has been developed which provides up to five sensors within the functional length of the urethra, addressing the simultaneity issue, and where each sensor has an axial symmetry, thereby addressing the circumferential sensitivity issue. Three sensors have been constructed and one of them has been subjected to approximately eight hours of use in a clinical setting during which 576 dynamic data points were obtained. The complete instrument system including the signal conditioning electronics, data acquisition unit, and the computer with its display and printer is shown in Figure 1. The technical details of the catheter sensor system itself is described in this paper.

### Transducer System Design

#### Requirements:

The requirement for the measurement system, as originally stated, was to be able to measure the closing forces along the continuous female urethra, to include an internal lumen for infusion or venting of the bladder, be very flexible and less than three-French in outside diameter, sense pressures in the range of 10 to 200 centimeters of water, and to be integrated into a computer system to facilitate the data reduction process. Continuous sensing of the closing forces within the functional length of the urethra was deemed impractical and was changed to "as many discrete measurement points as is practical." This translated into a cylindrical transducer structure having five sensing areas, each area approximately 0.2 inches long and within a length of one and one-half inches, approximately one-eighth inch in diameter, hollow, in order to contain an internal lumen, and a sixth sensing element located approximately one and one-half inches from the others. The resulting transducer system, without its protective sheath, is shown in Figure 2. Here one sees the five closely spaced sensor elements and a sixth element located some distance from the others to be always in the bladder and thus provide a reference. With this arrangement, discrete but simultaneous measurements are obtained throughout the urethra. Figure 3 is an enlarged view of one sensor element and is seen to be shaped like a "C". As pressure is applied around the circumference of this element displacement of the two halves (tending to close the longitudinal gap) results causing bending stresses to appear in the spine portion. These stresses cause minute deformations in the outer fibers of the metal which are then sensed with strain gages.

### Sensor Design

Each sensor element is cylindrical, nominally 0.120" diameter by 0.118" long by 0.004" thick; has a slot approximately 0.010" wide cut in the axial direction; and has slots machined transverse to the axis, leaving approximately 0.040" of circumferential area to act as a spine. To obtain a quantitative relationship for the sensor, first convert pressure in centimeters of water to pounds per square inch:

$$P_{\text{psi}} = 0.014215 \times P_{\text{cm-H}_2\text{O}}$$

Therefore:

$$P_{\text{min}} = 0.142 \text{ psi} \quad \text{and} \quad P_{\text{max}} = 2.843 \text{ psi}$$

This distributed pressure acting over the movable portion of the circumference can be represented as an equivalent force acting at the midpoint of each segment which then produces a bending moment at the root or spine. Young's modulus is then used to relate stress to strain as follows:

$F = P \times A$	$P = \text{Pressure and } A = \text{Area}$
$M = F \times L/2$	$L = \text{Length and } M = \text{Root Bending Moment}$
$s = \frac{Mc}{I}$	$s = \text{stress, } c = \text{distance to outer fiber}$
$\epsilon = s/E$	$I = \text{Moment of area of sensor segment}$
	$\epsilon = \text{strain, } E = \text{Young's Modulus}$

Combining the applicable constants and values gives the following expression for strain as a function of pressure tending to close or compress the sensor:

$$\epsilon = 2.378 \times 10^{-6} \times P_{\text{cm}}$$

This being the expression for mechanical strain at the root of the movable section, strain gages are placed in this area to convert the minute mechanical deformations into electrical signals proportional to strain. At the minimum pressures anticipated, the strains are very small, and every effort must be made to minimize extraneous effects and to enhance the signal at the point of inception. This is accomplished by placing a full, active four-arm Wheatstone bridge at the measurement site. In this manner, two gages are mounted in each half of the "C" section, one mounted in the circumferential direction or maximum strain direction, and the other mounted normal to the first. Connected in this fashion, two gages sense the compressive strains present on the inner surface; and two gages, mounted at 90 degrees to the others, measure the Poisson's ratio strains. Significantly, all gages see the same environment which minimizes the effects of extraneous inputs.

The above analysis must be qualified somewhat, acknowledging the finite size of the strain gages and the presence of an elastomeric sheath surrounding the final sensor. Finite size of the gages means that the entire gage does not see the level of strain as shown above, but rather an average over its sensitive area; and the sheath is a non-linear elastic member spanning the measuring section which effectively presents some load-carrying capacity in parallel with the measuring "C" sections. The non-linear effects are assumed small at this point, and the gage installation effects are accommodated through calibration.

The usable output voltage or signal from the strain gage bridge is dependent upon the gage factor, the number of active bridge arms, and the excitation or bridge voltage. The present configuration consists of two active arms (plus two Poisson's ratio arms) with a nominal gage factor of two. Semiconductor strain gages exhibit gage factors on the order of 120, sixty times that of the foil gages used, which would significantly improve the signal-to-noise ratio; however, semiconductor gages require a relatively flat surface because of their brittleness. In this application, the tight radius of curvature precludes their use. Utilizing a large, bridge excitation voltage will also maximize the output and signal-to-noise-ratio; however, one must always be cognizant of the self heating of the gage caused by  $I^2R$  losses which introduces long-term drift of the system. Combining all the applicable factors, one arrives at the following expression for sensor output voltage:



$$\text{Output Voltage} = 4 \times \text{Bridge Voltage} \times \epsilon$$

which for a bridge excitation voltage equal to 2 Volts,

$$\text{Output Voltage} = 19 \times 10^{-6} P_{cm}$$

From this it is seen that the range of output voltages from the transducer is from approximately 190 to 3800  $\mu$ volts. Most acquisition or display devices require signals in the  $\pm 2.5$  volt range. Thus it is obvious that a significant amount of amplification or gain in the system is required.

### Signal Conditioning

Signal conditioning consists of individual channel signal isolation, offset zero and gain adjustments, and low pass filtering of the data. The data is passed through isolation amplifiers and filtered to pass frequencies below 150 Hertz. The overall gain is adjustable and typically on the order of 250 to 300 making full scale signals approximately 1 Volt. The data is then passed to a sampling and analog-to-digital conversion subsystem where the data is quantized to 12-bit precision, and 250 samples from each channel are accumulated at a 100 samples-per-second rate. These records are then stored on disk for post test analysis.

### Transducer Calibration

A hydrostatic calibrator was constructed which consisted of a sealed chamber in which the catheter could be horizontally placed and water allowed to enter, expelling any air from the chamber. The chamber is connected via a flexible hose to an accumulator which can be raised or lowered to change the pressure in the chamber. Once the air is completely expelled from the system, the pressure is allowed to equilibrate with the level of water in the accumulator adjusted to the level of the catheter. At this point, the channels are adjusted for zero output after which the accumulator is raised to 100 centimeters and the gain adjusted to some convenient value. Iteration between the zero level and the 100cm level a few times and adjusting the zero and gain controls at each end will remove any interactions, and a convenient conversion factor, say 200 cm/volt, is established. By adjusting the height of the water column in a definite sequence, one can determine the drift, hysteresis, and linearity. Because pressure is a scalar quantity, it exerts the same influence in all directions making a good calibration of the instrument to measure circumferential closing forces.

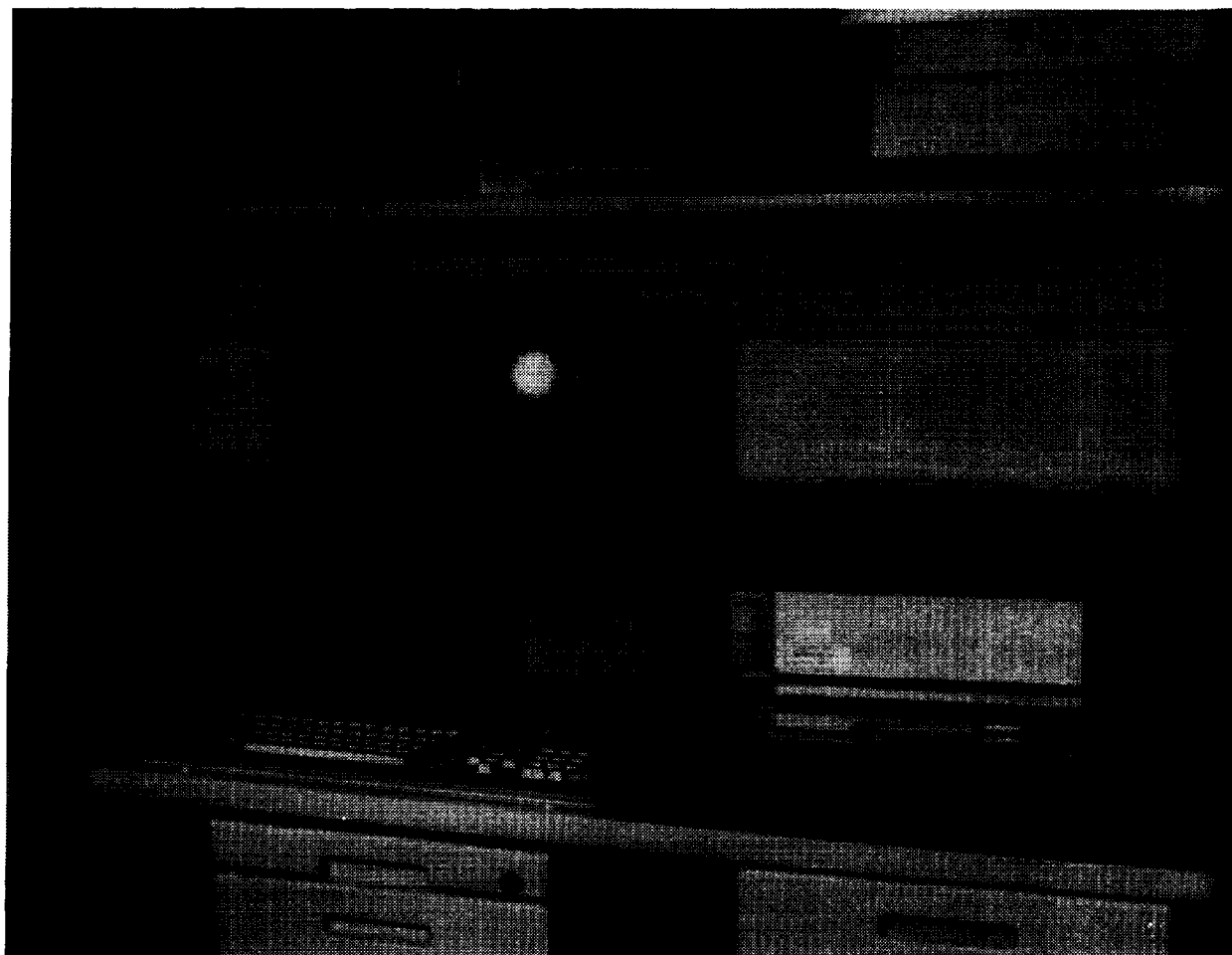
### System Output

As stated previously, the data is accumulated on a hard disk in the computer making it possible to perform any number of analysis procedures. Figures 4a and 4b illustrate two different formats that were used up to this point. Figure 4a includes test parameter and system information as well as a plot of the actual time history values, whereas figure 4b is merely a full-page representation of the particular time history. The two figures are from two different subjects, and one can see distinct differences in the shape of the time history.

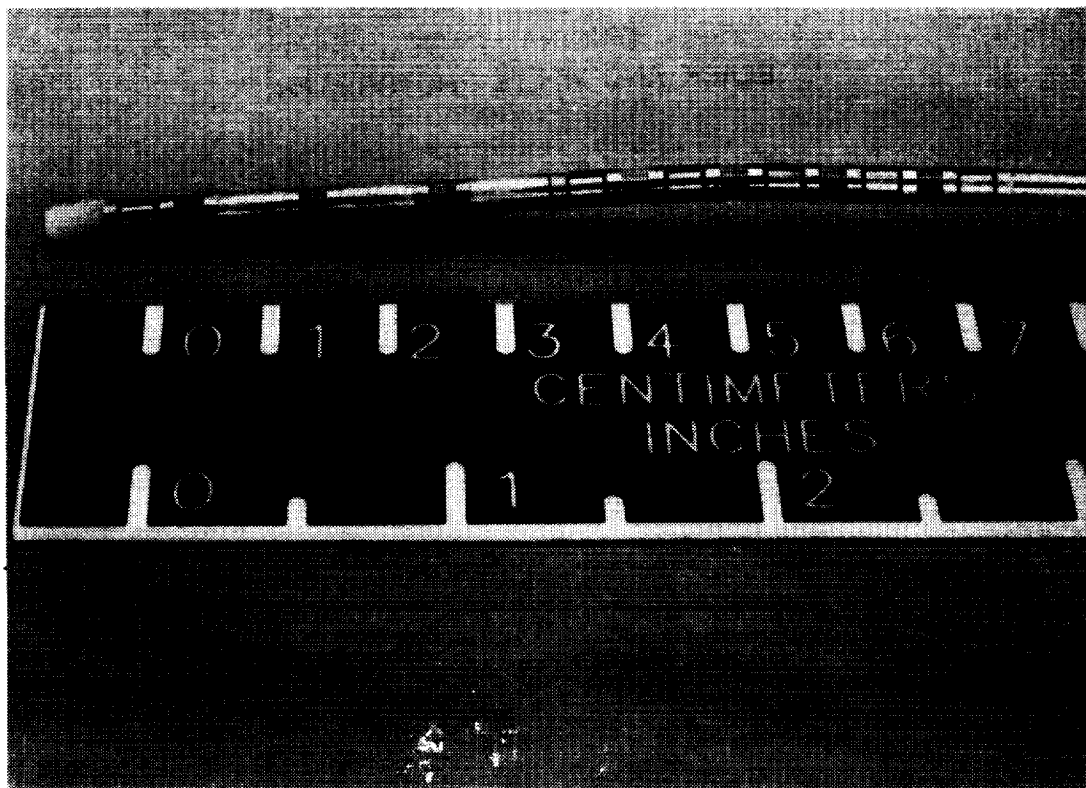
## CONCLUDING REMARKS

Based on the needs expressed at the outset, it is concluded that a sensor system has been developed which captures the essence of the requirements. Simultaneous measurement, at several discrete locations within the urethra, and insensitive, though not yet proven absolutely, to circumferential positioning of the catheter are features of the design. As a result of the clinical testing, a new catheter has been fabricated which is physically shorter and has the segments positioned closer together. This allows more sensor elements to be positioned within the effective length of the urethra, potentially four of the five as opposed to three for functionally shorter urethral lengths. Previously, the medical specialist was interested primarily in the amplitude at the respective element location and its ratio to the bladder pressure. With the new sensor and the associated data acquisition and display capability, not only amplitudes but also shape, rise and fall times, and muscular response propagation times are new parameters upon which diagnostic evaluations can be made.

ORIGINAL PAGE  
BLACK AND WHITE PHOTOGRAPH



**Figure 1: Signal Conditioning and Data Acquisition Instrumentation.**



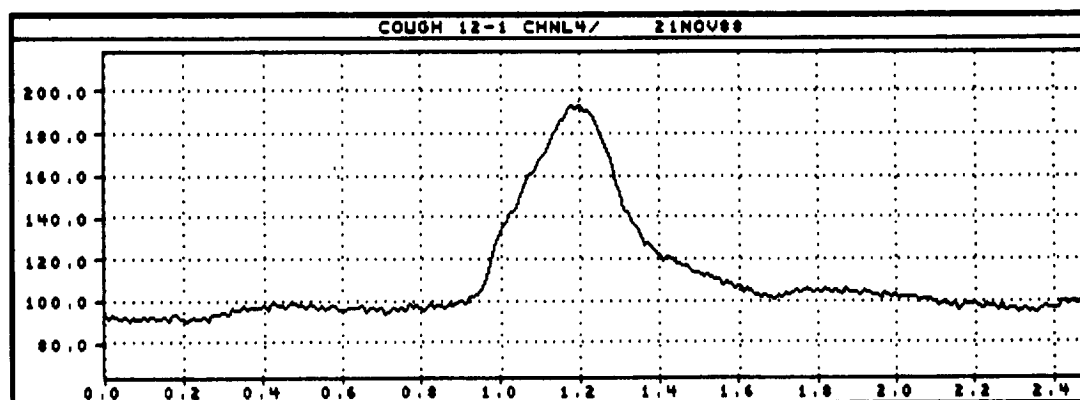
**Figure 2: Six Element Catheter without Sheath**



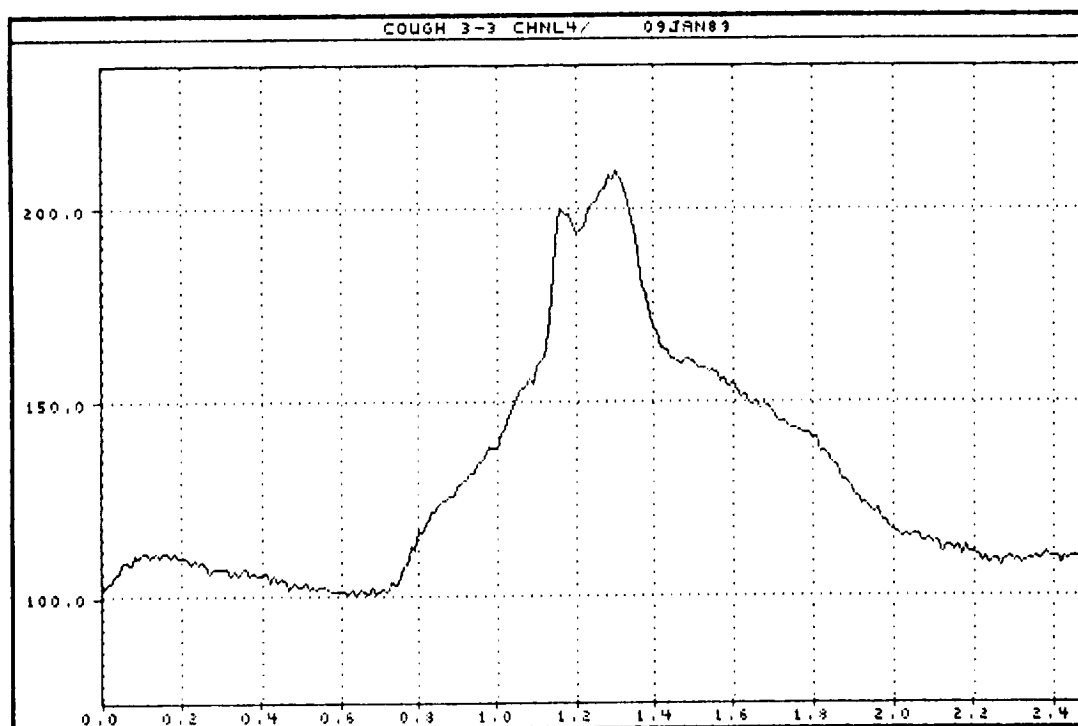
**Figure 3: Enlarged View of Single Element**

Datafile Name:	Unspecified
Version Number:	8
Signal Name:	Unspecified
Date Acquired:	11-23-1988
Time Acquired:	9:09:58.58
Vert Units:	Pressure
Horiz Units:	Secs
Num. Samples:	258

Sample Rate:	188
Maximum:	192.796882
Minimum:	89.987793
Comments:	Signal Generated in Analysis



**Figure 4a: Output Display, Format No. 1**



**Figure 4b: Output Display, Format No. 2**

## Electron Tunnel Sensor Technology

T.W. Kenny, S.B. Waltman, J.K. Reynolds and W.J. Kaiser

Center for Space Microelectronics Technology

Jet Propulsion Laboratory

California Institute of Technology

Pasadena, CA 91109

**Abstract**

We have designed and constructed a novel electron tunnel sensor which takes advantage of the mechanical properties of micro-machined silicon. For the first time, electrostatic forces are used to control the tunnel electrode separation, thereby avoiding the thermal drift and noise problems associated with piezoelectric actuators. The entire structure is composed of micro-machined silicon single crystals, including a folded cantilever spring and a tip. The application of this sensor to the development of a sensitive accelerometer is described.

Devices used in the measurements of many physical quantities (position, motion, pressure, radiation, temperature) rely on conversion of the physical quantity to a change in the relative separation between a pair of components. Changes in relative position are then detected through changes in the capacitance between the two elements,<sup>1-3</sup> through optical techniques,<sup>4,5</sup> or through flexing of piezoresistors.<sup>6</sup> In each case, large structures are often required to achieve sensitivities that are not limited by readout noise in the transducer. The availability of a more sensitive transducer would, in many cases, allow the construction of sensors with greater sensitivity, smaller mass, smaller volume, less power consumption, or greater bandwidth.

Recently, electron tunneling through a narrow vacuum barrier has been employed as a technique, Scanning Tunneling Microscopy (STM), for imaging the atomic scale structure of surfaces.<sup>7</sup> The tunneling current,  $I$ , has the following dependence on the separation between a pair of metallic electrodes,  $s$  :

$$I \propto V \exp(-\alpha \sqrt{\Phi} s), \quad (1)$$

where  $\Phi$  is the height of the tunneling barrier,  $V$  is the bias voltage,  $V$  is small compared to  $\Phi$ , and  $\alpha = 1.025 (\text{\AA}^{-1} \text{ eV}^{-1/2})$ .<sup>7</sup> For typical values of  $\Phi$  and  $s$ , the current varies by an order of magnitude for each 1  $\text{\AA}$  change in electrode separation. This sensitivity to relative position is superior to that available in other compact transducers. Since tunneling only occurs in regions where the tip is within several  $\text{\AA}$  of the surface, the active area of the sensor is microscopic. A position sensor based on electron tunneling has already been incorporated into the design for an accelerometer,<sup>8</sup> and several other applications are being considered.

Stable electron tunneling requires control of the separation of the tunnel electrodes to a small fraction of an Angstrom. An analog feedback circuit is used to compare the measured tunnel current to a set-point. The feedback circuit adjusts the voltage applied to an electromechanical actuator to correct for discrepancies between the measured tunnel current and the set-point. The bandwidth of the feedback circuit is limited to less than the resonant frequency of the actuator in order to avoid instability.

Typical vacuum tunneling devices rely upon piezoelectric electromechanical actuators. These actuators suffer from sensitivity to thermal drifts, hysteresis, and creep in the response of the piezoelectric materials; these effects impose severe limitations on the performance of existing tunnel devices. Also, efforts to miniaturize sensors which incorporate piezoelectrics are made difficult by the variety of different materials required for the construction of the device.

There have been recent advances in the sophistication of micro-machining of silicon through the use of anisotropic etchants as well as doping to control etching.<sup>9</sup> These advances have led to the development of a new class of sensors composed entirely of micro-machined silicon. There are several advantages to the development of sensor components in silicon. First, the well-developed techniques of photo-lithography may be used to control the structure of compact sensor components on the  $\mu\text{m}$ -scale. Also, silicon offers extremely high tensile strength combined with relatively low density. Further, silicon single crystal structures show highly reproducible characteristics. Finally, micro-machined silicon sensors are mass-produced inexpensively, and are easily miniaturized or modified to meet the requirements of a particular application.

We have designed and constructed an electron tunnel sensor composed of micro-machined silicon. Micro-machining has been used in this case to produce cm-scale

components with  $\mu\text{m}$ -scale precision. In contrast to all previous vacuum tunneling devices, the relative position of the electrodes is controlled through use of electrostatic forces applied between the elements. The electrostatic forces induce deflection of a micro-machined silicon cantilever spring. Replacement of the piezoelectric actuator with an electrostatic actuator is important for the following reasons. First, the electrostatic actuator is insensitive to thermal drifts and immune to the problems of creep associated with piezoelectric actuators. Also the response of the electrostatic actuator is a function only of the geometry and mechanical properties of the device, whereas the response of the piezoelectric actuator is also dependent on the characteristics of the piezoelectric material, which may not be reproducible between devices or over time. Finally, the electrostatic actuator may be miniaturized more easily because the scaling laws are known exactly, and the fabrication is less complex than for the piezoelectric actuator. The device described here can be modified for incorporation into a wide variety of sensors. Monolithic devices including sensor and control electronics are also feasible.

The tunnel sensor described here consists of three components which are constructed from  $200\ \mu\text{m}$  silicon wafers. The wafers have been polished on both surfaces and are coated with  $>0.5\ \mu\text{m}$   $\text{SiO}_2$  which is patterned by standard photo-lithographic techniques. The wafers are etched in Ethylene Diamine Pyrocatechol (EDP), removing the parts of the silicon wafers not covered by the  $\text{SiO}_2$  mask. After etching, the remaining oxide is removed in a buffered HF etch. A new oxide layer  $>1\ \mu\text{m}$  thick is then grown on all surfaces of the structure. Gold electrodes are thermally evaporated onto the components the sensor through shadow masks which have been fabricated by the same micro-machining techniques. The  $\text{SiO}_2$  serves as a dielectric isolation layer between the metal films and the silicon substrate. Figure 1 shows a sketch of the components which result from etching and metallization. The components are approximately  $4\ \text{cm}^2$  in area. The inner rectangular area of the folded cantilever spring can be deflected upward or downward relative to the outer segments by application of an electric potential between the large electrode and a corresponding deflection counter-electrode which is deposited on another component of the sensor. Given the mechanical properties of the silicon as well as the dimensions of the spring structure, we can calculate the properties of the spring. Because the spring constant scales rapidly with the dimensions of the legs, its characteristics can easily be tailored to meet the needs of a specific application.



Various methods for manufacturing a suitable tunneling tip are available.<sup>7,10</sup>

Recently, we have developed the following procedure for micro-machining a silicon tip directly from the silicon substrate. An oxide-coated surface of the silicon is first patterned by photo-lithography to leave only a 60  $\mu\text{m}$  x 60  $\mu\text{m}$  square of oxide. When etched in EDP, the edges of the oxide are undercut. When the undercutting is complete, the square fragment of oxide is carried away, leaving a pyramid-shaped silicon tip. A scanning electron microscope micrograph of a typical tip is shown in Fig. 1. The active surfaces of the tips are prepared by evaporation of 3000 Å thick Au films through a shadow mask. The tunnel sensor is then assembled as shown in Fig. 2.

Once assembled, a bias voltage is applied to the electrostatic deflection electrodes to close the electrodes and establish a tunnel current. Active regulation of the tip-electrode separation using feedback control of the tunneling current is carried out as for STM. The tunnel position sensor implemented in this way is generally applicable to a wide range of sensing applications. The properties of a tunnel sensor applied to measurements of acceleration are described below. Operation of the device as an accelerometer may be achieved in either of two ways. In each case, acceleration is measured by sensing the deflection of the spring-supported silicon mass.

In the first approach, denoted as open loop, acceleration is measured at frequencies above the feedback loop bandwidth. When the sensor is subjected to an acceleration,  $\partial a$ , the deflection of the folded cantilever spring,  $\partial z$ , is given by  $\partial z = \partial a / (\omega_0^2 - \omega^2)$ <sup>11</sup>, where  $\omega_0 = \sqrt{k/m}$  is the resonant frequency of the spring supported mass.<sup>12</sup> Since the current varies as  $\partial I = -I \alpha \sqrt{\Phi} \partial z$ , the responsivity is given by :

$$\frac{\partial I}{\partial a} = - \frac{I \alpha \sqrt{\Phi}}{(\omega_0^2 - \omega^2)} . \quad (2)$$

If the noise in the device is dominated by shot noise in the tunnel current,  $I_n = \sqrt{2 e I}$ , the noise-equivalent acceleration (NEa) is given by :

$$\text{NEa} = \frac{I_n}{\partial I / \partial a} = \frac{(\omega_0^2 - \omega^2)}{\alpha} \sqrt{\frac{2 e}{I \Phi}} . \quad (3)$$

The noise-equivalent acceleration is defined as the minimum acceleration that could be detected in a 1 Hz bandwidth.

As shown in (2), the responsivity in the open loop method varies as  $\sqrt{\Phi}$ , which may depend on operating conditions. Variations in  $\sqrt{\Phi}$  or other parameters which effect

the dependence of the tunnel current on electrode separation may be corrected for by adding a small modulation to the actuator voltage and measuring the resulting modulation in the tunnel current at that frequency. The ratio of the amplitudes of modulation directly yields the sensor sensitivity. This simple in-situ calibration may be performed continuously during operation of the sensor for applications which require high absolute accuracy.

In the second approach, denoted as closed loop, acceleration is measured for all frequencies less than the feedback loop bandwidth. In this case, the feedback loop responds to an acceleration by adjusting the force applied by the actuator so as to prevent any changes in the relative displacement of the tunneling electrodes. In this approach, the variations in the voltage applied to the actuator become the signal. For the electrostatic actuator in the present device, the force required to maintain constant electrode separation is given by :

$$F = m a = \frac{\epsilon A V^2}{2 d^2} , \quad (4)$$

where A is the area of the deflection electrodes and d is the separation between the electrodes. The responsivity is then given by

$$\frac{\partial V}{\partial a} = \frac{m d^2}{\epsilon A V} . \quad (5)$$

An advantage of this closed-loop approach is that the responsivity (5) is now independent of the height of the tunnel barrier or other properties of the tunnel electrodes as opposed to case for the open-loop approach. Shot noise in the tunnel current at frequencies within the bandwidth of the feedback loop causes fluctuations in the deflection voltage. These variations in deflection voltage produce changes in tunnel electrode separation, and therefore, variations in tunnel current which cancel the shot noise. The amplitude of the shot noise-induced voltage fluctuations is given by :

$$V_n = \sqrt{\frac{2 e}{I \Phi} \frac{k \epsilon A V}{m^2 d^2}} . \quad (6)$$

The noise equivalent acceleration sensitivity is obtained by dividing the responsivity (5) by the noise (6) and is exactly the same as the open loop sensitivity (3).

The sources of noise observed in vacuum tunneling devices are not fully understood. At high frequency, shot noise and Johnson noise are expected to dominate. At lower frequencies, noise with a 1/f characteristic typically dominates. This noise may originate from slow drifts in the device structure, or environmental sources of noise, such as vibration, or air currents. Thermal drift, which leads to differential thermal expansion may also contribute to the low frequency noise in vacuum tunneling devices. Other

mechanisms for  $1/f$  noise have been proposed.<sup>13,14</sup> With the replacement of the piezoelectric actuator by an electrostatic actuator, the problems associated with hysteresis, creep and thermal drift in piezoelectric materials should be greatly reduced. Contributions to the noise in tunneling due to fluctuations in the momentum transported have been considered.<sup>15,16</sup> This source of noise is not expected to be significant in our measurements.

The ultimate sensitivity of the tunnel sensor is limited by the noise characteristics of the vacuum tunnel junction. With the feedback loop closed and constant electrode separation being maintained, the tunnel current noise spectrum can be measured. Typical spectra for a current of 1.3 nA and bias of 100 mV are shown in Fig. 3 for different settings of the bandwidth of the feedback circuit. For this current, shot noise should contribute  $2 \times 10^{-14} \text{ A}/\sqrt{\text{Hz}}$ , which is below the observed noise. The current noise due to the components of the electrical feedback circuit is also well below the observed noise. The tunnel current noise power spectrum varies approximately as  $1/f$  for frequencies below  $10^3$  Hz. This result is consistent with noise measurements reported elsewhere,<sup>17</sup> and is likely due to environmental noise sources such as vibrations or thermal drifts. At frequencies greater than  $10^4$  Hz, an electrical resonance related to the deflection electrodes introduces additional noise.

By applying a low-frequency modulation to the tunnel current set-point in the feedback circuit, a modulation in the deflection voltage and tunnel current are produced. The barrier height,  $\Phi$ , may be determined from the ratio of the amplitudes of the modulations. For the Au electrodes in this device, measurements of  $\Phi$  have given a value of 0.5 eV. With  $\Phi = 0.5 \text{ eV}$ , the data in Fig. 4 implies that the measured sensitivity to variations in the tip-sample separation is  $3 \times 10^{-4} \text{ Å}/\sqrt{\text{Hz}}$  at a frequency of 1 kHz. Given a cantilever mass of 30 mg and a measured spring constant of 60 N/m as well as the dependence of the tunnel current upon separation, the responsivity of the sensor to changes in acceleration may be calculated. Below the resonant frequency of about 200 Hz, the responsivity is  $3 \times 10^{-4} \text{ A/g}$ . With the observed noise of the sensor, this corresponds to a sensitivity of  $1 \times 10^{-8} \text{ g}/\sqrt{\text{Hz}}$  at 10 Hz. At 1 kHz, the sensitivity is  $1 \times 10^{-7} \text{ g}/\sqrt{\text{Hz}}$ . In each case, the measured noise included contributions from accelerations in the laboratory, so the actual sensitivity is probably somewhat better.

achieved. This would allow low-cost production of sensors of this type as transducers for application to a wide variety of measurement needs. In particular, miniature pressure sensors and microphones could be constructed with tunnel sensor technology. The use of a tunnel sensor as the transducer in a pneumatic infrared detector is currently being explored.<sup>18</sup>

The research described in this paper was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory, California Institute of Technology, and was jointly sponsored by the Defense Advanced Research Projects Agency and the National Aeronautics and Space Administration, Office of Aeronautics, Exploration and Technology.

For comparison, the sensitivity of a capacitive displacement transducer may be calculated. It is assumed that the electrode separation is no less than 1/100 of the electrode length, that the voltage is no greater than the limit for dielectric breakdown in vacuum, and that sensitivity is limited by the shot noise at the ac measurement frequency. With these assumptions, the capacitive transducer has a theoretically-limited sensitivity to variations in electrode separation of approximately  $0.2 \text{ \AA}/\sqrt{\text{Hz}}$ . In this limit, the theoretical capacitive transducer is more than 3 orders of magnitude less sensitive than the existing tunneling transducer, and more than 4 orders of magnitude less sensitive than the theoretical limit to the tunneling transducer.

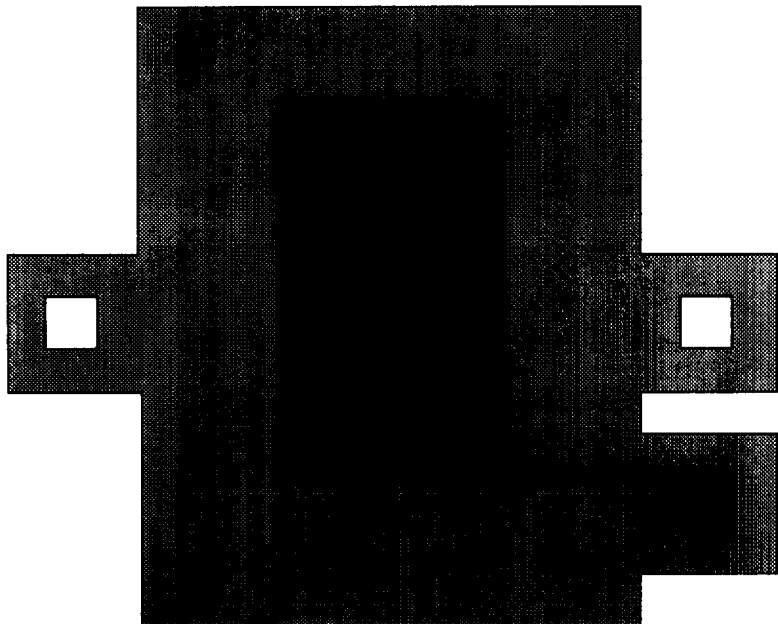
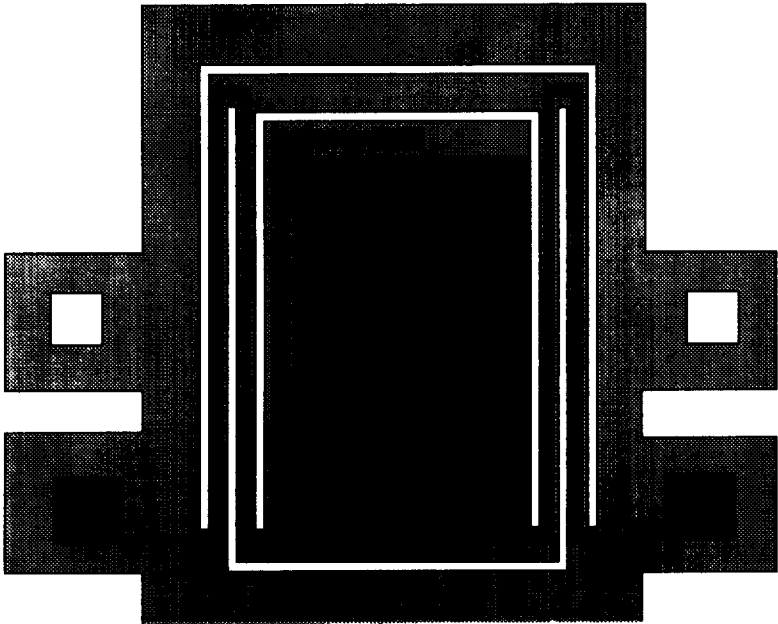
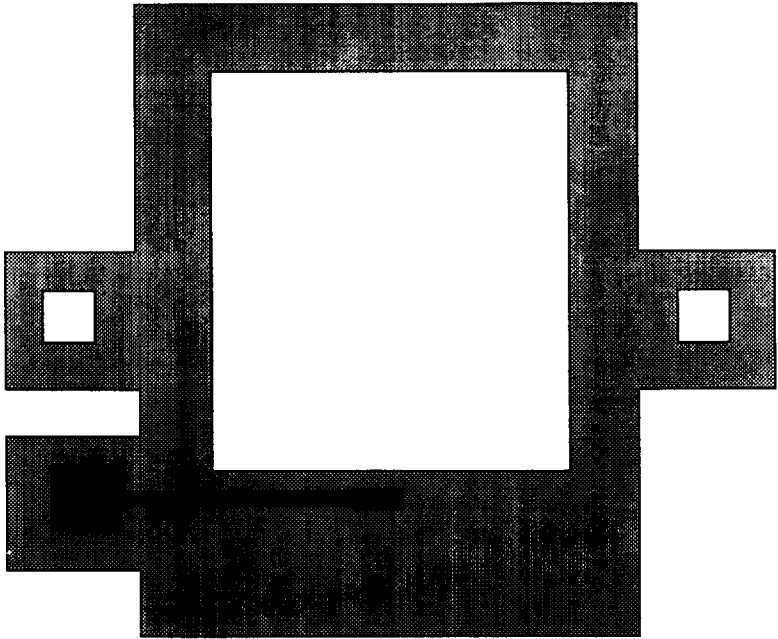
Since all of the components of the tunnel sensor can be prepared using standard lithographic techniques, miniaturization over several orders of magnitude can be readily

- <sup>1</sup> F. Rudolf, *Sensors and Actuators* 4, 191 (1983).
- <sup>2</sup> K.E. Petersen, A. Shartel and N. Raley, *IEEE Trans. Electron Devices* ED-29, 23 (1982).
- <sup>3</sup> C.S. Sander, J.W. Knutti, and J.D. Meindl, *IEEE Trans. Electron Devices* ED-27, 927 (1980).
- <sup>4</sup> E. Stemme and G. Stemme, *IEEE Trans. Elect. Dev.* 37, 648 (1990).
- <sup>5</sup> D.L. Gardner, T. Hofler, S.R. Baker, R.K. Yarber and S.L. Garrett, *J. Lightwave Tech.* LT-5, 953 (1987).
- <sup>6</sup> H.V. Allen, S.C Terry and D.W. De Bruin, *Sensors and Actuators* 20, 153 (1989).
- <sup>7</sup> G. Binnig and H. Rohrer, *IBM J. Res. Develop.* 30, 355 (1986).
- <sup>8</sup> S.B. Waltman, W.J. Kaiser, *Sensors and Actuators* 19, 201 (1989).
- <sup>9</sup> K.E. Petersen, *Proc. IEEE* 70, 420 (1982).
- <sup>10</sup> R.B. Marcus et al. *Appl. Phys. Lett.* 56, 236 (1990).
- <sup>11</sup> In the low-frequency limit, this reduces to  $\partial z = (m/k) \partial a$ , which is Hooke's Law. Damping has been ignored in this discussion.
- <sup>12</sup> It is assumed that the surface of the tunnel counter-electrode is flat and therefore  $ds = dz$ . If the surface of the counter-electrode is rough,  $ds = dz/\cos(Q)$ , where  $Q$  is the angle between the normal to the surface and the  $z$  axis.
- <sup>13</sup> M.E. Welland and R.H. Koch, *Appl. Phys. Lett.* 48, 724 (1986).
- <sup>14</sup> J.B. Pendry, P.D. Kirkman and E. Castano, *Phys. Rev. Lett.* 57, 2983 (1986).
- <sup>15</sup> M.F. Bocko, K.A. Stephensen, and R.H. Koch, *Phys. Rev. Lett.* 61, 726 (1988).
- <sup>16</sup> B. Yurke and G.P. Kochanski, *Phys. Rev. B* to be published (1990).
- <sup>17</sup> D.W. Abraham, C.C. Williams, and H.K. Wickramasinghe, *Appl. Phys. Lett.* 53, 1503 (1988).
- <sup>18</sup> T.W. Kenny, S.B. Waltman, J.K. Reynolds, and W.J. Kaiser to be published.

## FIGURE CAPTIONS

- Figure 1.** A drawing of the components of the tunnel sensor is shown. The folded cantilever spring with integral tip is shown on the lower right. It is manufactured by patterning both surfaces of the silicon wafer and etching until the tunnel tip had formed. The components which support the tunneling counter-electrode and the deflection counter-electrode are shown on the left and are manufactured by patterning only one surface of the wafer. Au films are deposited through micro-machined shadow masks to form the deflection and tunneling electrodes. The tabs with the square holes allow the components to be constrained laterally by alignment pins.
- Figure 2.** A SEM micrograph of a typical micro-machined silicon tip. The tip has been coated by 3000 Å of Au.
- Figure 3.** An exploded view of the assembled tunnel sensor showing the orientations of the components. From top to bottom are the deflection counter-electrode, the tunneling counter-electrode, and the folded cantilever spring. The shaded regions on the spring correspond to the Au films which are deposited through shadow masks.
- Figure 4.** This figure is a plot of the spectrum of the measured noise in the tunnel current at a bias of 100 mV and current of 1.3 nA while operating in air. The bandwidth of the feedback circuit was deliberately adjusted to be less than 0.1 Hz for the data shown as circles. For the data plotted as squares, the bandwidth was set at the maximum value that did not produce oscillations, which was 200 Hz for this device. A resonance in the electrostatic deflection system contributes to the noise at frequencies above 5 kHz.

FIGURE 1





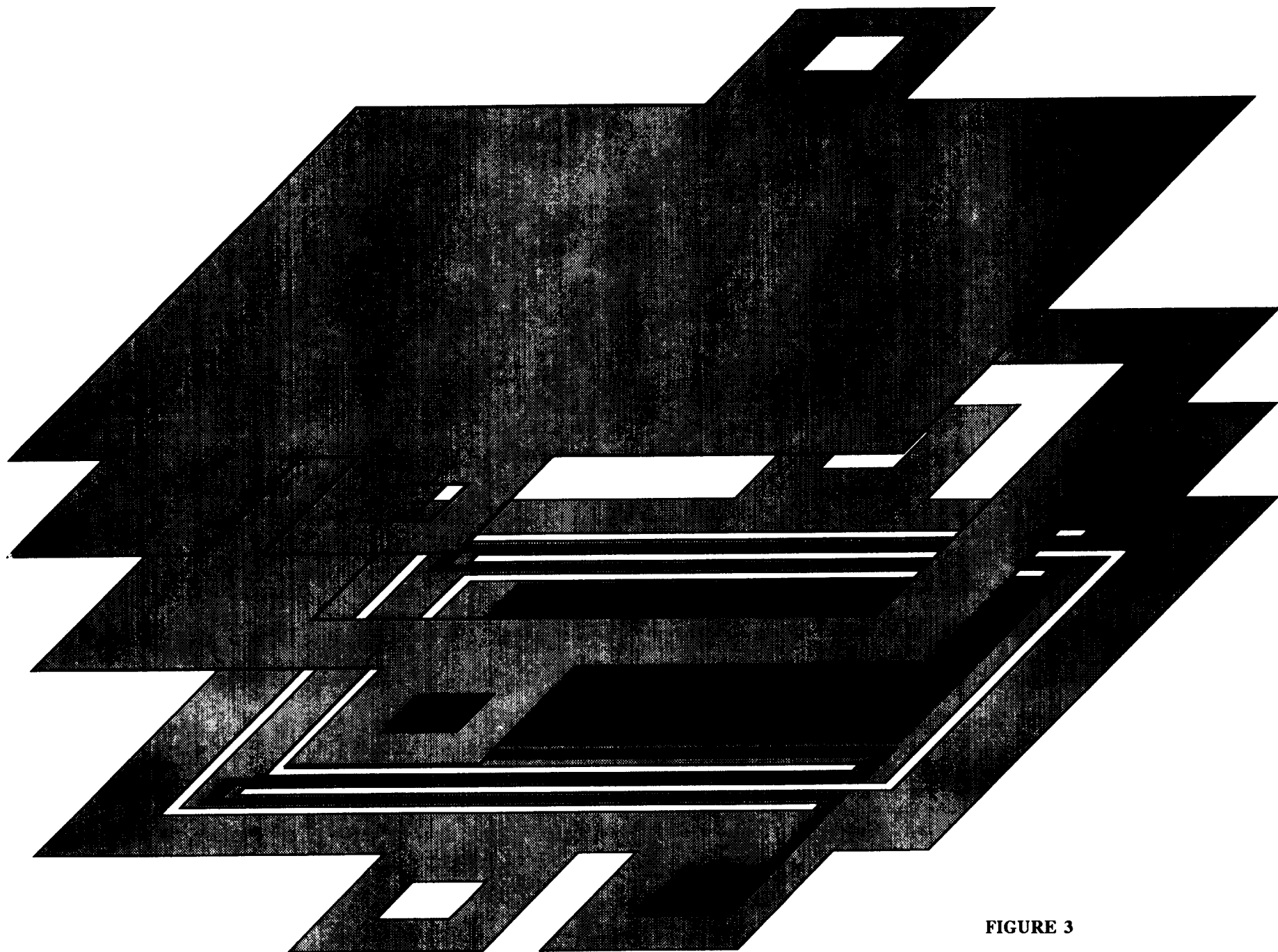


FIGURE 3



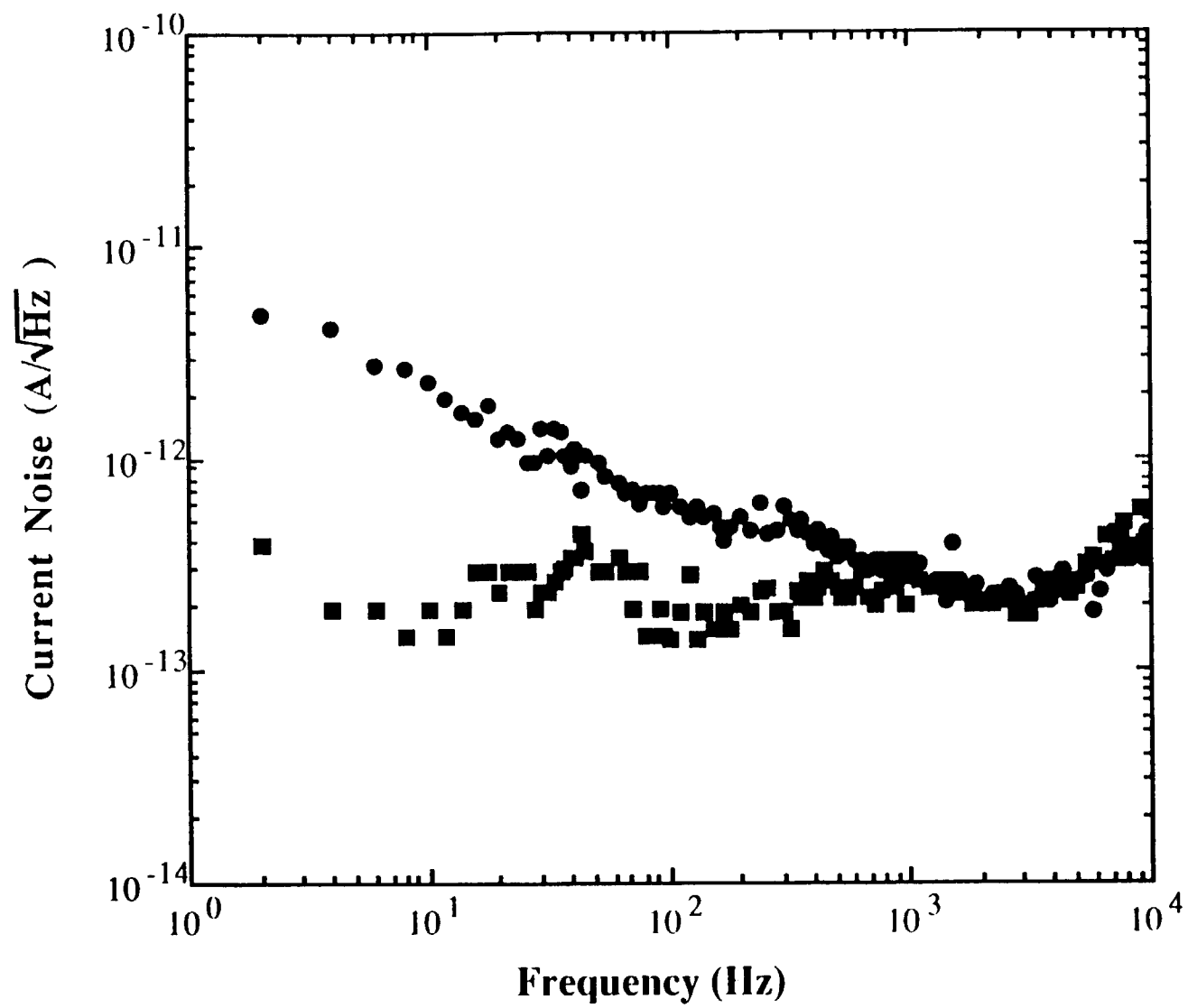


FIGURE 4

# PRACTICAL APPROACHES FOR APPLICATION OF RESISTANCE TYPE STRAIN GAGES ON HIGH TEMPERATURE COMPOSITES

Thomas C. Moore, Sr.  
NASA Langley Research Center  
Hampton, Virginia 23665

## INTRODUCTION

The introduction of carbon/carbon and metal matrix composites as most probable candidate materials for extensive use on hypersonic flight vehicles generates the need to measure their mechanical properties under severe test environments. Surface mounted strain gages have historically provided the most widely used means for measuring strains and still appear to be the most straightforward and least expensive method for making the strain measurements associated with these materials.

Because of the extreme temperatures associated with the utilization of these high temperature composites, two basic problems occur when attempting to instrument the composites with strain gages. One is the difficulty in reliably attaching strain gages to the surfaces of the unique materials; the other is the inaccuracy often associated with strain gages at extreme temperatures.

This paper will address four areas of interest with respect to utilizing strain gages on carbon/carbon (with SiC surfaces) and titanium matrix composites. The four areas are as follows:

- \* Strain gage and adhesive combinations on carbon/carbon (C/C) at temperatures from -190 degrees C (-310 degrees F) to 540 degrees C (1004 degrees F).
- \* Half-bridge gaging for reducing apparent strain on C/C using poisson's ratio and bending configurations.
- \* Review of the "field installation" techniques developed for gaging a C/C hypersonic generic elevon.
- \* Results of initial strain gaging efforts on titanium matrix composites.

Current research in developing techniques for increasing the maximum temperature for strain gages on carbon/carbon will be reviewed.

## STRAIN GAGE AND ADHESIVE COMBINATIONS FOR CARBON/CARBON

The primary difficulty in attaching strain gages to the surface of carbon/carbon is the arduous task of bonding the adhesive itself to the surface. To clarify this stated difficulty, it is important to note that, for this paper, all references to carbon/carbon surfaces include a surface layer of silicon carbide (SiC) encapsulating the carbon/carbon and it is the adherence to the silicon carbide that is difficult. This section of the paper discusses the techniques and materials developed for gaging carbon/carbon when testing at temperatures from -190 degrees C to 540 degrees C.

### Plasma Sprayed Strain Gage Installations w/Ceramic Precoat

Early attempts at strain gaging C/C for testing at temperatures above 370 degrees C (700 degrees F) involved the application of the gage with various ceramic cements. Six different ceramics were tested and all showed various degrees of micro-cracking which destroyed the gage. The failures with ceramic cements and oxygen/acetylene flame spraying were followed by the development of a gaging technique employing plasma arc flame spraying. Initial efforts had the entire installation performed using only the plasma gun in conjunction with Alumina powders. This technique worked but the loss rate during installation was high. In order to decrease the loss rate during installation and to improve strain data reliability, the procedure was modified and the detailed gaging steps are as follows: Mask the areas to be gaged and micro-sandblast them using 50 micron grit aluminum oxide abrasive. Next, apply a .001" thick precoat of Omega CC ceramic cement (modified by adding an equal amount, by volume, of SiC powder to the Omega base powder). After curing, plasma spray a .002" thick layer of Alumina over the precoat. The gage is then placed on the Alumina and a very thin coat of a ceramic cement (Hitec's Yellow Cerro works well) is applied over the gage convolutes to protect them from the plasma spray operation. The thin coat of ceramic cement is cured, and the strain gage is plasma sprayed with aluminum oxide. Several high temperature strain gage types were tested with this method and all survived the 540 degrees C. However, the foil type high temperature gages were more prone to gage failure during thermal cycling due to gage grid shearing. Later testing revealed that an oxygen-acetylene gun can be substituted for the plasma gun and the bond obtained is sufficient to read strains at 540 degrees C. However, the shear strength of the oxygen-acetylene bond is lower than that obtained with the plasma.

### High Temperature Gaging Using Ceramic Cement Only

While the plasma spray method of installing high temperature strain gages works well, there is another technique for providing strain gages on C/C for testing at 540 degrees C that does not require sophisticated gaging apparatus. It employs a ceramic cement which was not available when the plasma technique was developed. This ceramic cement supplied by Micro-Eng.II and called GC Ceramic does require careful application technique but when properly applied, a well bonded, reliable strain gage installation that is good to 540 degrees C and sometimes higher is achieved. With this ceramic, micro-cracking is greatly reduced and indiscernible at 540 degrees C. While data have been obtained at temperatures as high as 925 degrees C (1697 degrees F) with this cement it is not recommended above 540 degrees C on C/C with SiC surfaces because of the micro-cracking that occurs beyond this temperature. This ceramic works well with either wire or foil type high temperature strain gages. Following is a step-by-step procedure that has been utilized successfully with the GC ceramic cement on C/C specimens (w/SiC surfaces).

1. Micro-sandblast the areas to be strain gaged.
2. Apply a basecoat of GC ceramic cement to a thickness of .001 inches.
3. Air dry for a minimum of 30 minutes.
4. Slowly apply heat using a heat gun until the specimen temperature reaches 100 degrees C (212 degrees F) to 110 degrees C (230 degrees F) and hold for two minutes.
5. Cool to room temperature, inspect and then cure at 150 degrees C (300 degrees F) for one-half hour.
6. Cool and install the gage as follows:
  - a. place the gage on the basecoat and apply a minimum amount of GC ceramic in the open areas of the gage carrier frame. This coat should be sufficient to hold the gage convolutes in place, but thin enough to be able to still see most of the convolutes.

- b. cure the cement as in steps 3, 4, and 5. Cool to room temperature.
- c. Under a microscope, carefully remove the gage carrier frame and eliminate any loose particles or tape adhesive residue.
- d. fill all remaining open areas making certain that the cement "wets" these areas. This is important in preventing voids.
- 7. Cure as in steps 3, 4, and 5.
- 8. Next, another layer of GC is applied to fill any low areas, helping to provide a uniform ceramic surface.
- 9. Cure as in steps 3, 4, and 5.
- 10. After cooling, use a fine aluminum oxide stone to carefully hone any high areas.
- 11. Now, apply a final overcoat of the GC ceramic cement still maintaining a low ceramic profile.
- 12. Cure as in steps 3, 4, and 5.
- 13. Finally, take the specimen to 315 degrees C (600 degrees F) and hold for one-half hour.
- 14. Cool and microscopically inspect the installation. The entire installation should not be more than .008" to .010" thick.

#### HALF-BRIDGE GAGING FOR REDUCING APPARENT STRAIN ON CARBON/CARBON

While a fair degree of success has been realized in achieving a good bond for strain gages on C/C, the ability to accurately measure strains at temperatures above 370 degrees C (700 degrees F) is reduced by several factors. The steep slope of the apparent strain curve associated with high temperature strain gages (roughly 30 micro-strain/degree C w/BCL-3 gages on C/C), a degree of uncertainty in the repeatability of the apparent strain curve, the drift rate of the gage at high temperatures, and the often difficult task of accurately measuring the temperature, all contribute to the inaccuracies in high temperature strain measurements. Though half-bridge gaging configurations will be limited to certain applications, the data generated on C/C when half-bridges have been tested warrant serious consideration of this technique where possible. The BCL-3 strain gage has two features which make it desirable for half-bridge gaging. One, the scatter in total apparent strain from gage to gage is typically less than 500 micro-strain from a total of 17000 micro-strain at 540 degrees C, and, two, the variation in drift rate at high temperatures is negligible from gage to gage. Therefore, when a given test scenario will allow for two of these gages to be placed in adjacent legs of the bridge circuit with both gages being subjected to the temperature change, the resulting total apparent strain and drift rate are dramatically reduced. Figure 1 shows two typical apparent strain runs to 540 degrees C when half-bridge gaging is employed. One curve is generated with a bending half-bridge, the other with a poisson's ratio half-bridge. As shown, neither curve exceeded 250 micro-strain, yet, a single gage typically shows 17000 micro-strain at 540 degrees C on carbon/carbon with this type of gage.

#### Gage Selection For Testing At 540 Degrees C On Carbon/Carbon

Several gage types have been tested on C/C at 540 degrees C with two types out-performing the others. These are, the Batelle-Columbus-Laboratories gage type BCL-3 and the "Chinese Gage" distributed by Hitec Products and Micro-Engineering II. Both of these gages are manufactured from the same basic alloy and both have similar characteristics. One BCL-3 gage has been subjected to twenty temperature excursions to 540 degrees C on a C/C specimen with loads applied and is still

functioning. The "Chinese Gage" has been successfully tested on C/C specimens at 540 degrees C as part of a NASA, Langley Research Center and General Dynamics joint venture.

### "FIELD INSTALLATION" TECHNIQUES FOR GAGING A CARBON/CARBON ELEVON

A test program to be commenced this winter at NASA's Ames Dryden Flight Research Facility will require the installation of approximately 300 strain gages on an elevon constructed of carbon/carbon. This elevon is to be subjected to a series of load tests at room temperature followed by elevated temperature testing. Strain gage data at room temperature are essential and tentatively strain data will be obtained at elevated temperatures. A technique has been developed that will allow the elevon to be tested at a temperature of 200 degrees C (390 degrees F) as well as room temperature using conventional gaging materials. The technique, simple and straightforward, is described here. Also, later in the program, an attempt will be made to install high temperature gages using a ceramic cement for testing at 540 degrees C. That gaging technique is also reviewed here.

### Selection Of Strain Gages And Gaging Materials For Testing To 200 Degrees C

This program requires strain gaging for testing at room temperature and at elevated temperatures if possible. Since this will be a fairly lengthy program in terms of time required for testing, it was decided to utilize an adhesive that has proven itself to be excellent for long term reliability. Hopefully, this adhesive could be applied to the surface and the gage in a manner consistent with techniques applicable for structural component gaging. Several adhesives were tested in conjunction with Micro-Measurements, Inc. encapsulated series WK-00, WK-03, and WK-06 gages. The first two adhesive types tested bonded well but the gage failed during repeated excursions to 200 degrees C. These adhesives were Micro-Measurements, Inc. M-BOND GA-2 and M-BOND GA-61. The third adhesive, M-BOND AE-15, worked well initially and gage type WK-00-500BH-350 had generated the smallest apparent strains in early tests. Consequently, a C/C specimen was instrumented with one of these gages on each side placed back to back. The specimen was then subjected to ten apparent strain runs to 200 degrees C. Apparent strains were repeatable for both gages on all of the runs, gage output drift at 200 degrees C was zero after two hours, and bending strain data at 200 degrees C was repeatable within two percent. Because of the ease of installation and the success with the M-BOND AE-15 in conjunction with the WK-00-500BH-350 strain gage, no further effort was made to find another combination of gage and adhesive for the temperature range of 25 degrees C to 200 degrees C.

### Gaging Procedure Followed For Testing Carbon/Carbon At 200 Degrees C

In order to simulate, to a degree, the gaging scenario that will be required for the carbon/carbon elevon gaging, the gages were pre-wired and installed without the benefit of a microscope which is typically used during the installation. Following are the steps performed in gaging the C/C specimens discussed above.

1. Micro-sandblast areas to be gaged using 50 micron grit aluminum oxide. Also, sandblast spots for bonding lead wires, including thermocouples.

2. Using clean, dry air, blow away sandblasting residue.
3. In areas where gages are to be attached, apply a coat of M-BOND AE-15. This "precoat" should be thick enough to allow for sanding (after curing) in order to provide a relatively flat surface for the strain gage.
4. Slowly raise the temperature to 65 degrees C (150 degrees F) and hold for four hours. Use heat lamps or vacuum pad heaters to achieve the required cure temperature.
5. Following this cure, sand the AE-15 using a sanding block and 180 grit sandpaper. Sand away the precoat until the peaks of the SiC are reached. Remove the residue with clean air or pressurized freon.
6. Install the pre-wired gages using "Rumble Strain Gage Vacuum Pads" as per standard vacuum pad gage installation procedures.
7. Cure the installation for four hours at 65 degrees C as in Step 4.

#### Tentative Gaging Procedure For A Carbon/Carbon Elevon For Testing At Temperatures To 540 Degrees C

While the entire testing program for the elevon has not yet been finalized, a gaging procedure has been developed which will provide some degree of strain measurement at temperatures of 540 degrees C and perhaps higher. The procedure, though not described in detail here, is similar to that described earlier in the paper under the sub-heading "High Temperature Gaging Using Ceramic Cement Only".

#### INITIAL GAGING EFFORTS ON TITANIUM MATRIX COMPOSITES

This area of research has just recently been initiated at NASA's Langley Research Center as part of a joint effort with McDonnell Douglas, St. Louis, MO. While this effort includes research in developing techniques and strain gaging materials that will perform accurately and reliably at temperatures up to 815 degrees C (1500 degrees F) on metal matrix composites, the initial thrust of the research was in the cryogenic region.

#### Gaging And Testing Of Titanium Matrix Composite Specimen

Initial testing of a titanium matrix composite specimen, type: SCS-6/BETA 21S [0/90/0] involved strain gaging the specimen utilizing standard NASA-LaRC gaging procedures with conventional foil strain gages and gaging materials. The specimen was subjected to temperatures from -190 degrees C (-310 degrees F) to 260 degrees C (500 degrees F). The purpose of the test was to obtain baseline data at room temperature, at -190 degrees C, and at 260 degrees C. Two single gages of the same type were installed back-to-back with the only difference being in the self-temperature-compensation (STC) number of each gage. One gage was a WK-03-250BG-350 and the other was a WK-06-250BG-350. Cantilever bending loads were applied to the specimen at 25 degrees C, -185 degrees C, and at 265 degrees C. The strain variation at those temperatures is shown for each gage in Figure 2. Gage factor change has been accounted for at each temperature and, as shown, each gage generated identical strain levels at its respective temperature. Apparent strain runs were made through the temperature range of -190 degrees C to 265 degrees C. A typical run is shown in Figure 3 with outputs recorded at 50 degree C intervals. Note that the WK-06 gage generates less strain cryogenically but drops off dramatically above room temperature. The WK-03 gage while generating -1400 micro-strain at -190 degrees C is relatively flat from 25 degrees C to 265 degrees C.

Apparent strains were not corrected for change in gage factor as a function of temperature change. Research and development efforts are ongoing in this area, specifically, gages and bonding material combinations that will function at 815 degrees C.

#### CURRENT EFFORTS IN GAGING CARBON/CARBON FOR USE AT 815 DEGREES C AND HIGHER

Even though high temperature strain gages have been successfully attached to C/C (with SiC surfaces) and survived to temperatures of 1100 degrees C (2012 degrees F), there are problems to be overcome at these very high temperatures. These include, survival of the gage when thermal cycling is involved, resistance leakage to ground, and repeatability of apparent strain curves. Current efforts are being directed toward developing a technique that will provide a reliable strain gage on C/C at 815 degrees C. Good initial results are being obtained utilizing two new techniques and each is briefly discussed here.

#### Carbon/Carbon Gaged And Tested At 815 Degrees C Utilizing Plasma And Oxygen-Acetylene Powder Gun

A means of obtaining a good bond between the gaging ceramic and the silicon carbide surface has been obtained utilizing a combination of plasma arc flame spraying and oxygen-acetylene flame spraying. A combination of ceramic spray powders was also used. The basecoat was  $3\text{Al}_2\text{O}_3\text{-}2\text{SiO}_2$  (Mulite powder), plasma sprayed. This was followed by  $\text{Al}_2\text{O}_3$  (Alumina powder) which was oxygen-acetylene flame sprayed and was utilized to secure the gage to the basecoat of Mulite. The gage was then oversprayed with more Mulite using the plasma gun. Importantly, the bond was enhanced by first sandblasting the surface with a very coarse (24 grit) aluminum oxide abrasive. This abrasive, while effective in roughening the surface, may cause microscopic damage to the substrate. This damage possibility is currently being investigated at Langley.

Cantilever bending strain data, repeatable within four percent, were obtained at 815 degrees C using this method in combination with a Battelle-Columbus-Laboratories type: BCL-3 strain gage. Though this installation functioned well for several cycles at 815 degrees C, it failed during the second excursion to 930 degrees C (1706 degrees F). An inspection revealed micro-cracking of the Mulite.

#### Strain Data Obtained At 1100 Degrees Celsius On A Carbon/Carbon Specimen

A carbon/carbon specimen furnished by General Dynamics, Fort Worth, TX, was strain gaged and tested at temperatures up to 1100 degrees C. This specimen differed from all others tested to date, in that it had a very rough chemically vapor deposited (cvd) coating of silicon carbide added to the existing silicon carbide surface. The gaging approach was similar to that described above with a few exceptions. Half-bridge gaging was utilized to reduce apparent strain and Alumina was the only spray powder used. Figure 4 plots the bending strains at 1100 degrees C obtained during two excursions to that temperature. Unfortunately, the specimen did not have a surface sealant and since the tests were conducted in air the specimen delaminated after two runs to 1100 degrees C. Figure 5 is a photo of the specimen after the second excursion to the loadings test temperature. The delamination is visible throughout the length of the specimen. The strain gages are still functional.

## CONCLUDING REMARKS

Strain gaging any material for testing at extreme temperatures, whether its at -190 degrees C or at 815 degrees C, requires the utilization of the appropriate strain gages, gaging materials, installation accessories, and gaging configuration in order to provide the maximum accuracy and reliability for a given test requirement. The task of determining the optimum means and materials for strain gaging high temperature composites such as carbon/carbon and titanium matrix composites is particularly difficult. The maximum test temperature, the maximum strain level, type of test (dynamic, static, both) are but a few of the factors that will help in determining the gaging approach. The strain gaging of these composites for cryogenic and elevated temperature testing is relatively new. As these materials are made available to the strain gage community, improved approaches and techniques will be forthcoming.

While reliable strain gage installations on C/C have been demonstrated at temperatures from -190 degrees C to 540 degrees C and while current research indicates that strain data at 815 degrees C is readily achievable, the desired goal of routinely measuring strain at 1100 degrees C on C/C using surface mounted strain gages is still not obtainable.

## ACKNOWLEDGEMENT

The author wishes to express his appreciation to Mr. Fred Lamm (senior strain gage technician, Modern Machine & Tool Co. Inc.) for his art, skill, patience, and assistance throughout this effort.

## REFERENCE

1. M.M. Lemco, Characterization of BCL , Chinese, and Dentronics High Temperature Strain Gages. Paper number 42, Eighth National Aero-Space Plane Technology Symposium
2. S.J. Lanius, R.G. Brasfield, S.P. Wnuk, Development of a Strain/Temperature Gage and Attachment System for use on Carbon-Composites at Elevated Temperature, 4th Annual Hostile Environments and High Temperature Measurements Conference, SEM
3. J.W. Sawyer, T.C. Moore, Testing Issues for Carbon/Carbon Structures, NASA Conference Publication 3065-Vol.1
4. T.C. Moore, High-Temperature Strain Gage Installation Techniques for Carbon/Carbon With Silicon Carbide Surfaces, NASP Technical Memorandum 1113



# BCL GAGES

## Apparent Strain Curves

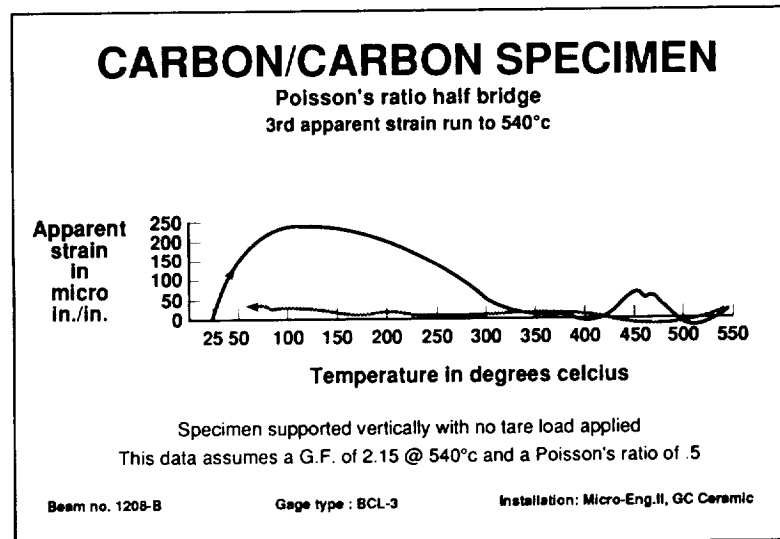
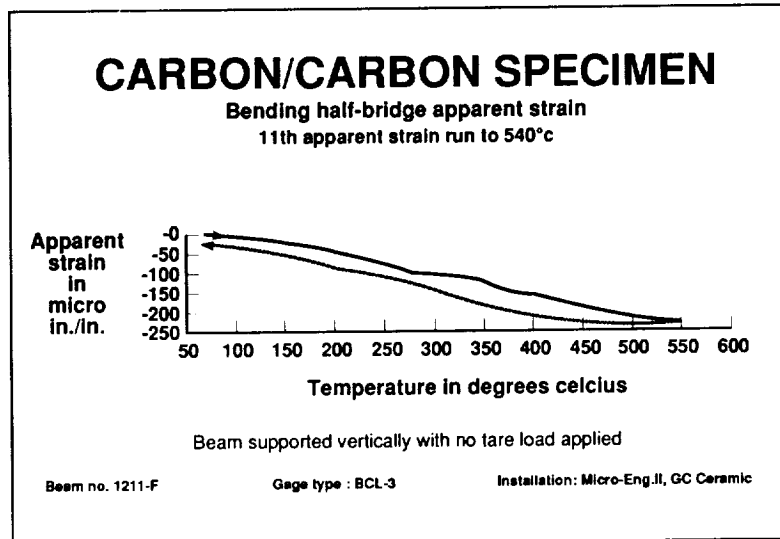
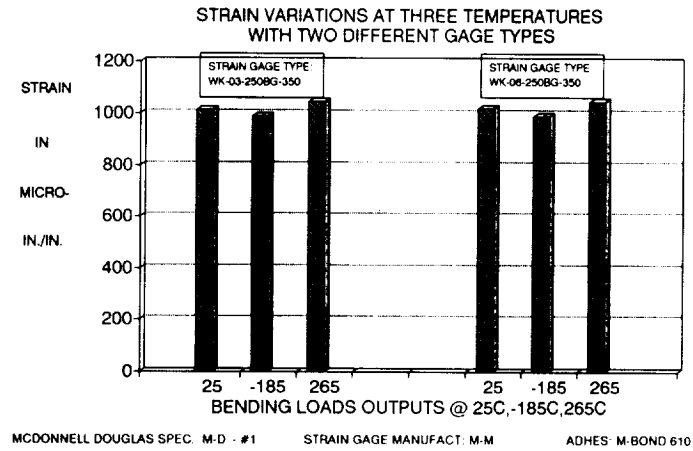
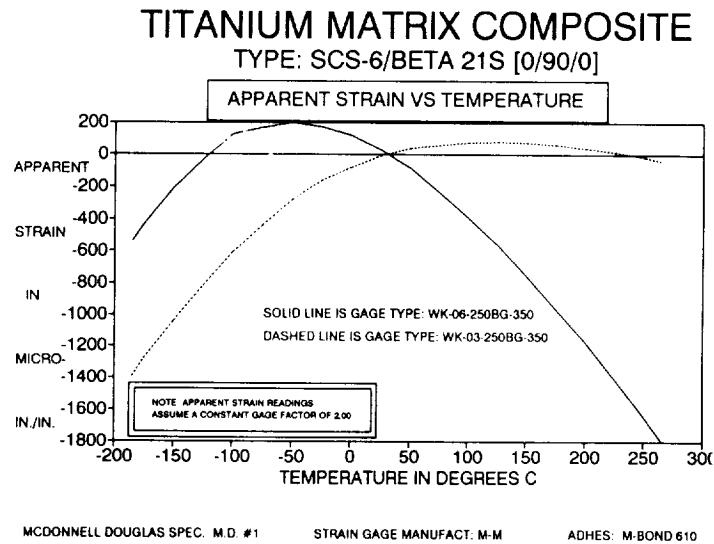


Figure 1

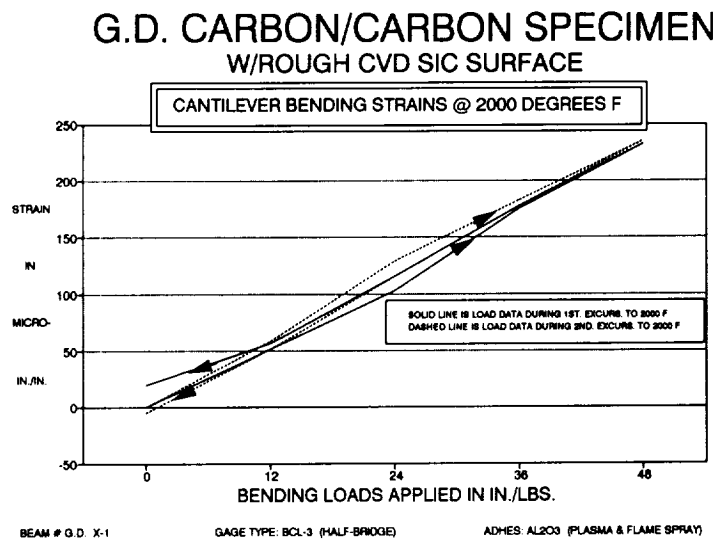
## TITANIUM MATRIX COMPOSITE TYPE: SCS-6/BETA 21S [0/90/0]



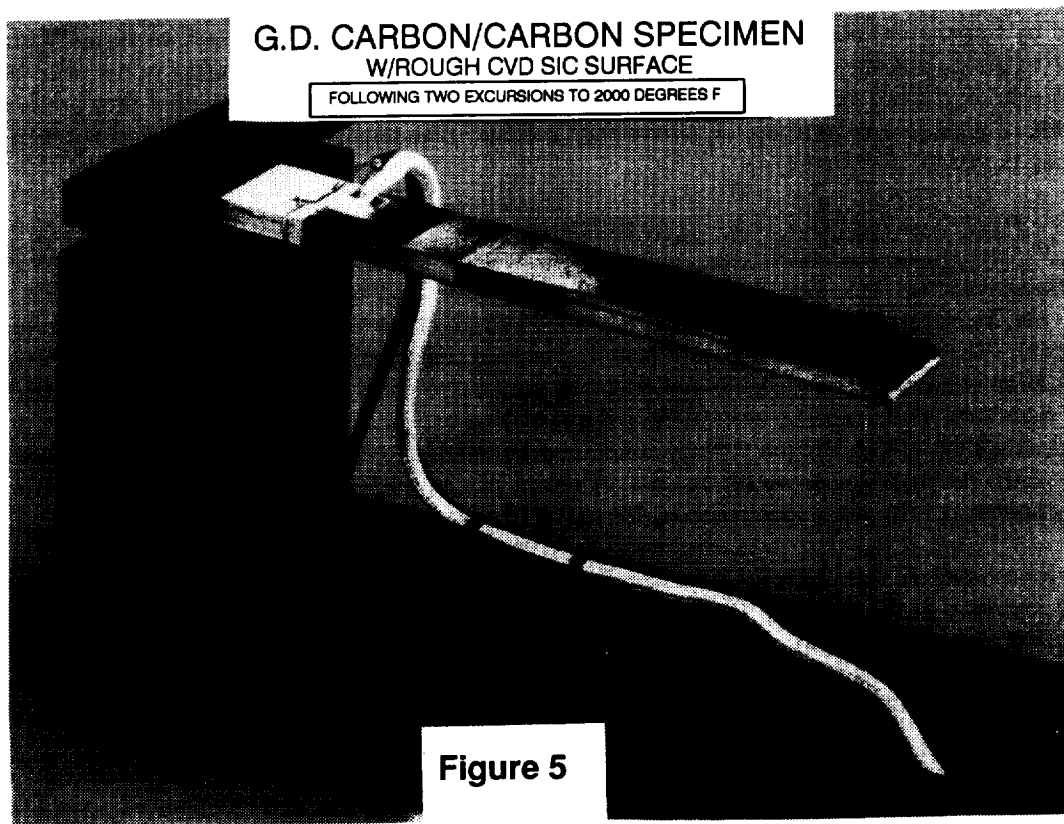
**Figure 2**



**Figure 3**



**Figure 4**



**Figure 5**

**ORIGINAL PAGE IS  
OF POOR QUALITY**

N91-28069

AI MASS SPECTROMETERS  
FOR SPACE SHUTTLE HEALTH MONITORING

F. W. Adams

NASA  
Mail Code DL-ESS-24  
Kennedy Space Center, Florida

BACKGROUND

Modern rocket development in the United States formally began in 1914 with Dr. Robert Goddard's patents on the concept of utilizing liquid nitrous oxide for the oxidizer and gasoline for fuel in self-contained, change of momentum based, main propulsion systems (1). His first flight, with hardware using gasoline and liquid oxygen, did not occur until 1926, some 12 years later. German acquisition of the concept prior to World War II led to Dr. von Braun's development of the V-2, powered by liquid oxygen and alcohol. Post war U.S. development of the Redstone, Jupiter, Atlas, Apollo, Shuttle, and ultimately, Advanced Launch System (ALS) are the historical background for this presentation.

Development of the concept of using liquid hydrogen for fuel began in 1960 with the beginning of the Apollo program. Fuels used prior to this were hydrocarbons (alcohol, kerosene, jet fuel), or solids. There is approximately a 700:1 volume reduction achieved by converting gaseous hydrogen at room temperature to liquid at cryogenic temperatures. The increase in specific impulse for hydrogen/oxygen over hydrocarbon/oxygen fuels is approximately 1.5:1. (2) The choice of hydrogen for fuel imposed greatly increased hardware requirements for being able to identify and evaluate leaks.

Shuttle is the first system where liquid hydrogen was used in the liftoff stage of the propulsion system. It is boosted by detachable solid rocket motors to help attain orbit. The launch countdown formally begins 3 days prior to liftoff, and the loading of cryogenic hydrogen and oxygen begins about 11 hours prior to liftoff. During launch countdown, ground support systems are prepared for use by verifying that they are functional and capable of holding a calibration. They are then carefully maintained "powered up", and watched for problems as the count continues.

INTRODUCTION

Mass spectrometers were first used in a ground support role in 1964 for Saturn I launch vehicle testing, for identifying hydrogen leakage in the main propulsion system of the S-IV second stage, during cryogenic operation (3). Early instruments for Saturn centered around a magnetic sector type analyzer. Shuttle instrumentation is based on quadrupole and magnetic sector type mass spectrometers. Other types of hydrogen detectors, while capable of locating and quantifying hydrogen leaks at low levels, are not usable above approximately 10% hydrogen concentration, and are unable to differentiate oxygen leaks from air intrusion, which the mass spectrometer can do. A recent patent, #4,953,976 based on Raman scattering, is the nearest technology to the mass spectrometer, for use in this type of service.

Mass spectrometers can be made to be sensitive to helium, the purge gas used in liquid hydrogen systems. Helium is also used as the trace gas (for safety reasons, during testing, when hydrogen is not present) as an aid in flagging and locating leaks. The general utility of the ruggedized mass spectrometer as a process instrument, as opposed to a laboratory curiosity, has led to its' increased use in facility ground support equipment to enhance the visibility of flight hardware with respect to liquid hydrogen and oxygen leakage.

Facility mass spectrometers were not installed for the classical liquid oxygen / hydrocarbon main propulsion system ground support equipment (GSE). In these systems, mass spectrometers existed in the form of drag-on helium leak detectors, using helium as a trace gas. They were not used to sense fuel leaks

directly. Permanent, facility analyzers, were installed for Shuttle because it is powered by liquid hydrogen, which produces leaks in real time, that cannot be detected by use of classical helium leak detectors.

### **MASS SPECTROMETER INSTALLATIONS AT KSC**

The facility Hazardous Gas Detection System (HGDS) at Kennedy Space Center, Florida (KSC), is a mass spectrometer based gas analyzer (4). Two instruments make up the HGDS, which is installed in a prime/backup arrangement, with the option of using both analyzers on the same sample line, or on two different lines simultaneously. It is used for monitoring Shuttle during fuel loading, countdown, and drainback, if necessary. The system is located in the mobile launch platform, underneath the vehicle, and must sustain the shock and vibration of launch. It must be operational during main engine firing for flight readiness firings (FRF), and after main engine firing, for launch, in order to perform during de-tanking, if launch is aborted.

HGDS is controlled and monitored, during countdown, from the Launch Control Center approximately four miles away. It was HGDS that discovered the hydrogen leaks in STS-6 Flight Readiness Firing (FRF), and STS-35 and STS-38 launch attempts, while monitoring the aft compartment of the orbiter.

Four compartments on the flight vehicle are monitored for their fraction of hydrogen, helium, nitrogen, oxygen, and argon. These are the External Tank/Inter-tank area, Payload Bay, Mid-body and Aft compartment. One sample is drawn from the hydrogen tail service mast (TSM). The presence of anything but nitrogen indicates main propulsion system, payload, fuel cell, or ground support hardware system leakage respectively. Launch Commit Criteria (LCC's), are determined prior to tanking, for the amount of each gas that will be allowed without taking action in some form.

Samples from the various compartments are pulled down dedicated stainless steel lines by small diaphragm type pumps. A sample switching arrangement, utilizing remotely controlled valves, is used to direct the sample to the analyzer. In order to reduce the time required for the sample from a particular area to reach the analyzer, a system of differential-flow / pressure-reduction stages are used. Disruption of flow in the sample delivery system can cause the analyzer to report data on a non-representative sample, and thereby give a distorted picture of the status of the vehicle.

Additional mass spectrometers have been added, as facility hardware, to monitor the 17 inch hydrogen orbiter/ET disconnect for flight readiness firings, special tanking tests, and launch. Plans are in process to add mass spectrometer capability to monitor the hydrogen fill and drain lines on the ground side, to more quickly determine the state of purge before and after fueling operations. A miniature, ruggedized flight version of the facility mass spectrometers is in the prototype stage at this time, and may some day be flight hardware. Mass spectrometers have flown in space aboard satellites and planetary probes since the beginning of the program.

### **THE NEED FOR ARTIFICIAL INTELLIGENCE (AI)**

The use of complex instruments, operated over many shifts, by different people, in a flexible operational environment, has caused problems in tracking status of GSE and the vehicle. A requirement for overall system reliability has been a major force in the development of Shuttle GSE, and is the ultimate driver in the choice to pursue AI techniques for Shuttle and ALS mass spectrometer systems. The need for certainty that the numbers displayed on the operators' console are truly representative of the variables that are being measured, is critical.

### **SHUTTLE APPLICATIONS OF AI**

There are five areas that have been identified, where AI techniques can help solve KSC's problems for Shuttle. First is the pre-launch system validation. It takes about two weeks to validate HGDS for launch. This involves a complete check of critical system voltages, gain and offset adjustments in critical stages,

verification of the amount of consumables (calibration gases), redundant circuitry checks (filaments in the ionizer, vacuum gages, etc.), and verification of normal operation, both locally, and from the remote control stations in the firing rooms. Many parts of this operation can be automated. Use of an Advisor here would be to guide personnel through the instrument checkout procedures.

Second, is the reduction of display complexity, as seen by the operators and users, during operations. This is a serious problem when more than one mass spectrometer display must be viewed simultaneously. Often two instruments are used to look at the same sample to verify that the data is consistent. An ideal display driver would indicate only the gas concentrations on the sample lines and would have the mass spectrometer function completely transparent to the operator. An Advisor here would provide visibility of background functions when problems occur, as a console operator aid. Near real time operation here is a firm requirement. Figure 1 shows the current display for the HGDS console for Shuttle. Figure 2 shows the actual variables (function designators) displayed, and where they appear on the screen.

Third, is the requirement to verify that samples from the vehicle are within the allowed limits, LCC'S, which normally vary, depending on the state of loading. During countdown, console operators are constantly looking at the data for clues that there might be something wrong with it. The sample seen by the analyzer must be representative of the environment around the hardware being evaluated for leaks. There must be no unknown system hardware failures which might distort the readings seen by the operators. An Advisor here would comment on the certainty of the data, including cross correlation between other instruments that might be looking at the same gas stream. Historical data from earlier launches could be referred to if deemed significant. Near real time performance is required at this phase.

LCC's are sometimes real-time variables, depending on the situation, and must be constantly observed in this light. LCC's for hydrogen, oxygen, and helium concentration, apply to ground and flight hardware during cryo loading sequences up to liftoff, or drainback and boiloff, if launch is not possible. The vehicle data is monitored by many knowledgeable persons at different locations on KSC and at other NASA centers (JSC, MSFC), and it is important, for this reason, that this data be representative and reliable. System oversight by knowledgeable instrument operators is necessary because data users will not normally be aware of abnormal system operation.

The fourth requirement is that real time troubleshooting of one of these systems (not to mention three or four), is a major undertaking, particularly when being performed simultaneously with launch support. KSC Design Engineering personnel, in coordination with operations, are developing techniques and special test sequences that will improve the console operators' ability to perform advanced troubleshooting during countdown, to verify whether a problem is in the vehicle or the analyzer, and evaluate the impact on the launch operation. As a side note, the first thing the data users question is not their system, but the integrity of data that indicates hardware problems. The advisor here would suggest testing routines that could be run real time while the system is still gathering data in support of the launch countdown.

The fifth and last application of AI is a tool for training new operators and upgrading skills of those familiar with the instruments. One serious problem with complex hardware is the limited number of people who understand specific instruments. During high pressure situations, where testing of a vehicle on one pad is being worked simultaneously with either testing or launching on another, skilled personnel are in short supply. Often the skills required for launch and testing are different, and less skilled people can be effectively used in testing situations. Allowing less skilled personnel to operate the system quickly can be a tremendous benefit to operations, provided everyone knows their limits and responsibilities. Here the advisor would track progress and advise operators where they need additional training to become proficient.

## CONCLUSION

It should be emphasized here that a specific instrument or capability will be installed for Shuttle only because the launch system cannot reliably function without it. AI would not normally be applied to a hazard warning system at KSC if it did not offer a significant advantage over manual methods of tracking problems

and reducing data in real time.

The burden of operating multiple mass spectrometers has increased greatly with each new system installation. The proper response to the situation is not to bring in more people, as there are only limited facilities in the firing rooms (console screen space, personnel access, etc.). ALS GSE will be deployed with similar motivation.

Shuttle has been used as a testbed for conceptual aspects of mass spectrometer application development for ALS. One reason that teams at KSC were ready to support special testing, when the latest group of leaks was discovered on Shuttle, was that personnel were looking at advanced applications of mass spectrometer systems in preparation for a demonstration of hardware for ALS.

In support of Shuttle and development for ALS GSE, it is important to keep the target in mind when designing a system of this magnitude. One is constantly being distracted by peripheral operations and special test support that are going on in parallel with system development. Often, new concepts are identified in these situations, and it is tempting to try to incorporate everything learned into new GSE.

Real system requirements must be kept isolated and not changed during system development. Experience with Shuttle has indicated that certain failure modes are more likely to occur than others. Equipment design must be based on knowledge of the hardware (what it is, how it is used, how it has failed in the past), and what kinds of surprises were seen when previous systems were installed. The ability to zero in on requirements, up front, is the key to achieving high and successful launch rates, in present and future systems.

#### BIBLIOGRAPHY

1. Robert Goddard, Pioneer of Space Research, Dacapo Series of Space, Dacapo Press, Inc., Subsidiary of Plenum Publishing Corp., 1963.
2. Starr, Stan, Personal conversation, 10/02/90.
3. Helms, W. R., History, Design and Performance of the Space Shuttle Hazardous Gas Detection System, Space Shuttle Technical Conference, NASA Conference Publication 2342, Part 1, 1983.
4. Helms, W. R. & Raby, B. A. A prototype Gas Detection System for NASA's Space Shuttle. 26th Annual Conference on Mass Spectrometry and Allied Topics, 1983.

## **APPENDIX A: SOFTWARE USED IN SYSTEM DEVELOPMENT**

**GOLDWORKS: (GOLDHILL COMPUTERS, INC., EXPERT SYSTEM INTEGRATED DEVELOPMENT ENVIRONMENT.)**

**LISP DEVELOPMENT ENVIRONMENT FOR GENERAL PROTOTYPING AND PRELIMINARY MODELING**

**CXPERT: SOFTWARE PLUS, LTD, CROFTON, MD.**

**EXPERT SYSTEM DEVELOPMENT ENVIRONMENT COMPATIBLE WITH THE GOAL OF ROMMING THE EXPERT SYSTEM VIA DOS OR UNIX. RULES, DATA BASE, INFERENCE ENGINE.**

**OS-9: MICROWARE CORP., DES MOINES, IOWA**

**ROMMABLE OPERATING SYSTEM AND DEVELOPMENT ENVIRONMENT**

**ZORTECH C+ +: ZORTECH CORP., WOBURN, MASS**

**AUXILIARY DEVELOPMENT, COMMUNICATION AND GRAPHICS INTERFACE, CXPERT SUPPORT, ETC.**

### **SALIENT HARDWARE FEATURES**

**ROMMED OPERATING SYSTEM AND APPLICATIONS FOR OPERATION IN A HIGH SHOCK AND VIBRATION ENVIRONMENT.**



## **APPENDIX B: HARDWARE USED**

COMPAQ 386/LISP AND GRAPHICS TERMINAL

68030/TARGET EMBEDDED PROCESSOR TO BE USED IN THE LAUNCH ENVIRONMENT  
(FORCE/VME, GESPAC)

GMX / 68020 GMX CORP., CHICAGO, IL

MGA-1200 MASS SPECTROMETER, PERKIN ELMER CORPORATION

## **APPENDIX C: GODDARD'S PATENTS RELATIVE TO ROCKET PROPULSION DEVELOPMENT**

**FIRST EVIDENCE OF RUSSIAN INTEREST IN ROCKETRY 1881**

**TSIOLKOVSKY, INVESTIGATION OF SPACE BY MEANS OF ROCKETS 1903**

- |    |       |   |            |
|----|-------|---|------------|
| 11 | 02653 | MULTI STAGE ROCKET  | 07/07/1914 |
| 11 | 03503 | BREECH BLOCK TO INJECT SUCCESSIVE CHARGES, COMBUSTION CHAMBER, NITROUS OXIDE, ETC | 07/14/1914 |

**GODDARD, A METHOD OF REACHING EXTREME ALTITUDES 1919**

**GODDARD AWARE THAT GERMANY WAS INTERESTED IN ROCKETRY ALSO 1923**

**OBERTH FIRST PUBLISHED 1924**

- |         |  |            |
|---------|--|------------|
| 1879186 | IGNITION OF LIQUID FUEL, CURTAIN COOLING | 09/27/1934 |
| 1879187 | GYROSCOPIC STEERING                      | 09/27/1934 |

#### APPENDIX A: SOFTWARE USED IN SYSTEM DEVELOPMENT

**GOLDWORKS:** (GOLDHILL COMPUTERS, INC., EXPERT SYSTEM  
INTEGRATED DEVELOPMENT ENVIRONMENT.)

LISP DEVELOPMENT ENVIRONMENT FOR GENERAL PROTOTYPING  
AND PRELIMINARY MODELING

**CXPERT:** SOFTWARE PLUS, LTD, CROFTON, MD.

EXPERT SYSTEM DEVELOPMENT ENVIRONMENT COMPATIBLE WITH  
THE GOAL OF ROMMING THE EXPERT SYSTEM VIA DOS OR UNIX.  
RULES, DATA BASE, INFERENCE ENGINE.

**OS-9:** MICROWARE CORP., DES MOINES, IOWA

ROMMABLE OPERATING SYSTEM AND DEVELOPMENT  
ENVIRONMENT

**ZORTECH C++:** ZORTECH CORP., WOBURN, MASS

AUXILIARY DEVELOPMENT, COMMUNICATION AND GRAPHICS  
INTERFACE, CXPERT SUPPORT, ETC.

#### SALIENT HARDWARE FEATURES

ROMMED OPERATING SYSTEM AND APPLICATIONS FOR OPERATION  
IN A HIGH SHOCK AND VIBRATION ENVIRONMENT.

#### APPENDIX B: HARDWARE USED

COMPAQ 386/LISP AND GRAPHICS TERMINAL

68030/TARGET EMBEDDED PROCESSOR TO BE USED IN THE LAUNCH  
ENVIRONMENT (FORCE/VME, GESPAC)

GMX / 68020 GMX CORP., CHICAGO, IL

MGA-1200 MASS SPECTROMETER, PERKIN ELMER CORPORATION

#### APPENDIX C: GODDARD'S PATENTS RELATIVE TO ROCKET PROPULSION DEVELOPMENT

	FIRST EVIDENCE OF RUSSIAN INTEREST IN ROCKETRY	1881
	TSIOLKOVSKY, INVESTIGATION OF SPACE BY MEANS OF ROCKETS	1903
1102653	MULTI STAGE ROCKET	07/07/1914
1103503	BREECH BLOCK TO INJECT SUCCESSIVE CHARGES, COMBUSTION CHAMBER, NITROUS OXIDE, ETC	07/14/1914
	GODDARD, A METHOD OF REACHING EXTREME ALTITUDES	1919
	GODDARD AWARE THAT GERMANY WAS INTERESTED IN ROCKETRY ALSO	1923
	OBERTH FIRST PUBLISHED	1924
1879188	IGNITION OF LIQUID FUEL, CURTAIN COOLING	09/27/1934
1879187	GYROSCOPIC STEERING	09/27/1934

## INSTRUMENTATION FOR OPTICAL OCEAN REMOTE SENSING

W. E. Esaias  
Code 971, Ocean and Ice Branch  
Goddard Space Flight Center

### ABSTRACT

Instruments used in ocean color remote sensing algorithm development, validation, and data acquisition with potential for further commercial development and marketing will be discussed.

The Ocean Data Acquisition System (ODAS) is an aircraft-borne radiometer system, suitable for light aircraft, has applications for rapid measurement of chlorophyll pigment concentrations along the flight line. The instrument package includes a three channel radiometer system for upwelled radiance, an infrared temperature sensor, a three channel downwelling irradiance sensor, and Loran-C navigation. Data are stored on a PC aboard and processed to transects or interpolated "images" on the ground. The instrument has been in operational use for two and a half years, most extensively (over 35 missions) with the NOAA Chesapeake Bay Program, but also in NASA missions in the North Atlantic, Iceland, and off Brazil. The accuracy of pigment concentrations from the instrument is quite good even in complex Chesapeake Bay waters.

To help meet the requirement for validation of future satellite missions, a prototype air-deployable drifting buoy for measurement of near-surface upwelled radiance in multiple channels is currently undergoing test deployment. The optical drifter burst samples radiance, stores and processes the data, and uses the Argos system as a data link. Studies are underway to explore the limits to useful lifetime with respect to power and fouling. There is also high potential for development of simple, highly reliable sensors for optical radiance measurements from underway vessels.

### INTRODUCTION

The marine phytoplankton play an important role in the Earth's carbon cycle, and recent technology has provided the means to observe the marine biosphere in surface layers at global and monthly periods. Absorbance of light by chlorophyll pigments involved in photosynthesis causes a shift in peak reflectance from blue to green as the amount of phytoplankton increase which is readily detected by satellite ocean color sensors. Application of satellite visible remote sensing is essential to understand the spatial and temporal distributions of the marine biosphere, but is only one part of a complete observing system. Measurements from aircraft, ships, and in-situ observations are also crucial to understand the cycles of marine phytoplankton, the marine food web, and the role of the marine biosphere in global carbon and climate cycles. Development of supporting technology within the commercial sector is crucial to overcome the tremendous sampling challenge presented by the oceans. This paper describes two systems under development at Goddard which fill critical sampling needs for quantifying the satellite observations and which are felt to have significant commercial potential.

## THE OCEAN DATA ACQUISITION SYSTEM (ODAS): APPLICATION TO CHESAPEAKE BAY

NASA research has shown the utility of remotely sensed radiance for measuring chlorophyll plant pigment concentration for oceanic research programs. Between the demise of the Coastal Zone Color Scanner in June, 1986, and the flight of a follow-on sensor in the early 1990's, scientists are dependent upon aircraft sensors for contemporary data. Aircraft sensors also typically provide higher spatial and temporal coverage than satellite sensors, and so will always be an important complement to ship, buoy, and satellite observation systems. While several research ocean color instruments are available within federal laboratories, there is currently no simple, commercially-available instrumentation available for routine scientific use. This section describes the development and application of a nadir viewing radiometer system developed for monitoring chlorophyll pigments.

One approach to measuring phytoplankton biomass on the time and space scales of importance is remote sensing from aircraft and satellite to determine spectral properties that are responsive to pigment concentrations (Esaias, 1980). Remote sensing of chlorophyll pigment concentrations using satellite and aircraft sensors has matured during the last decade, but this technology has principally been applied to the open ocean where phytoplankton dominate the optical signal. Less work has been done in optically more complex waters typical of estuaries, principally because algorithms for deriving pigment concentrations there are very dependent upon other water optical properties. Since the relative optical characteristics vary greatly from one estuary to another, and with time, local algorithms must be developed and validated for optimal results.

Extensive use of aircraft ocean color instruments has been made by NASA and NOAA scientists (Campbell & Thomas 1981; Hoge & Swift 1981; Campbell & Esaias 1983, 1985; Campbell et al. 1986; Hoge et al. 1986, 1987). These studies encompass a variety of water types, including some Case 2 waters (coastal, estuarine) and both "active" and "passive" systems. Active systems measure laser-induced fluorescence of chlorophyll; passive systems measure upwelled radiance. Many of the existing systems, especially active systems, are quite large and/or have requirements necessitating the use of a large aircraft. Costs are prohibitive for programs such as Biomonitoring that require frequent and regular sampling.

### System Description

The Ocean Data Acquisition System is a low-cost instrument with potential commercial application. It is easily installed in small aircraft and flown over the coastal zone ocean to remotely measure sea surface temperature and three channels of ocean color information. From this data, chlorophyll levels can be derived for use by ocean scientists, fisheries, and environmental management. Data can be transmitted to ships for real-time use with other measurement and sampling objectives.

The optical portion of the system has three primary instruments: an IR radiometer to measure surface temperature, a three-channel visible spectro-radiometer viewing the water, and a three channel sky sensor mounted on top of the aircraft to monitor incident irradiance. The down looking radiometer channels have a field of view of 2 degrees, and measure upwelled radiance at 460, 490, and 520 nanometers (blue, blue-green, and greenish blue) in 15 nm bands. The outputs are fed into a PC based data system where they are digitized to 12-bit resolution, formatted, and recorded on a hard disk. All parameters, including spectral curvature used for chlorophyll calculation, are displayed on the monitor. The operator controls the external shutters, gain, and starts and stops data recording from the keyboard. Radiometric data are sampled every 0.1 sec, which at typical flight speeds, corresponds to sampling distances of from

5-10 meters. The aircraft package contains a LORAN-C unit for aircraft location information (updated every 8 seconds), on-board data processor and formatter, digital cassette tape recorder, . The initial storage of data is as binary files uniquely designated by the time of their creation. A typical file from the applications discussed includes a record of 2 to 15 minutes duration corresponding to an individual flight line, with smaller files being produced in cross-estuary tracks and larger files in along-axis tracks.

The data telemetry system is based on a packet radio terminal controller, and radio transceiver for data transmission to a ship. From the measurement flight altitude of 500 ft., the line of sight transmitter range to a ship is about 30 miles. The VHF transceiver can also be used for voice communication and coordination between aircraft and ship. Digital data rates are about 1200 baud.

The shipboard package contains a transceiver packet terminal controller, data processing capability, cassette tape recorder, and printer. Both raw data and chlorophyll concentrations are available for real-time analysis, and these data can be shown in latitude and longitude coordinates on a screen map. The shipboard package has not been used as much as initially planned, since the emphasis to date has been on monitoring rather than support of real-time experiments.

To keep down the cost of reproducing the instrument system, commercially available subsystem components are principally used. Standard camera lenses for the optical systems, amateur radio packet controllers, and modified amateur transceivers (modified to FCC assigned ODAS frequencies) for the data transmission systems, and rack mountable personal computers for the data systems are examples. The detector assembly was designed at GSFC. The detectors, filters, and preamplifiers are in a temperature-controlled housing to maintain radiometric stability. The sky sensors are modified from commercially available cosine collectors. A thermistor is use to monitor the sky sensors to correct for their temperature response.

Data processing on the ground has been performed on PC's, MacIntosh, and Sun computers. The principle functions are 1) calculation of navigation coordinates from the 8 second Loran data to match the more frequent radiometric data, 2) eliminating individual points contaminated by sun glitter through screening based on sample variance, 3) application of calibration and pigment algorithms, 4) merging the navigation with averaged, glint free data, 5) contouring and displaying the data using various packaged programs.

Two algorithms have been used to process the radiance data from ODAS. The system was designed to exploit the "curvature" algorithm discovered by Grew (1981) at NASA's Langley Research Center. The curvature algorithm:

$$\log(\text{Chl}) = a - b \log [(L_2)^2 / (L_1 * L_3)]$$

where  $L_n$  represents the radiance detected in channel  $n$ , and  $a$  and  $b$  are system coefficients, is very amenable to the ODAS objectives, since it minimizes the need for atmospheric correction and is less sensitive to variations in other water constituents than other approaches (Campbell and Esaias, 1983). The coefficients  $a$  and  $b$  are initially derived from comparisons with in-situ data, but thereafter ODAS does not depend upon the availability of sea-truth. Since the conversion of digital counts to radiance is a linear function, the algorithm can be applied to raw data (using slightly different coefficients).

Secondly, algorithms using ratios of channels can also be used provided attention is paid to the variations in optical properties other than in-water pigment concentration. Development of band-ratio algorithms for estuarine systems would be greatly enhanced by the addition of spectral bands corresponding to those planned for future ocean color satellites.

Ocean color algorithms, and curvature algorithms in particular, require high precision radiometry, and can tolerate very little drift between channels as a function of time. Repeated calibration using the GSFC hemisphere radiance source has shown precision at the 0.2% level between missions and channels. Absolute radiometric accuracy of the system is estimated at 7%, primarily a function of the accuracy of the hemisphere source. The signal to noise ratio is greater 2400:1.

### System Test and Evaluation

The ODAS system was tested on the Wallops P3-A Orion during development on a series of 4 missions including the Arctic Ocean Expedition in May 1987. An example of comparisons of ODAS observations with simultaneous NASA WFF Airborne Oceanographic Lidar values are given as Fig. 1. The correlation coefficient ( $r$ ) for the chlorophyll signals in this example was 0.97. These data were obtained between the Gulf Stream and the coast off Wallops Island, Va. in March 1988. This high degree of correlation is typical for the performance of curvature algorithms in ocean waters, however further testing was required to determine how well the ODAS and curvature algorithms performed in more turbid estuarine waters.

In July and September, 1988, the ODAS was used by a group of investigators from the Virginia Institute of Marine Sciences, The Chesapeake Bay Institute, University of Delaware, and NOAA to overfly portions of the Chesapeake and Delaware Bays (Harding, 1988). The objective of these flights were to test ODAS performance in the more turbid estuarine waters, to demonstrate ODAS utility as an operational sensor in the Chesapeake Bay monitoring program, and to demonstrate a transfer of technical capability from NASA to academic investigators. The ODAS was flown on the VIMS aircraft, a DeHaviland Beaver, and was operated by the investigators. About 25 megabytes of data from the flights were analyzed by scientists from the three institutions.

Based upon the success demonstrated in the initial program, Drs. L. W. Harding and E. C. Istweire of The Johns Hopkins Chesapeake Bay Institute, with funding from the NOAA Coastwatch Program, led a team of investigators in conducting a more ambitious series of flights in 1989 and 1990 (Harding and Istweire, 1990). The purpose of these flights was to monitor phytoplankton dynamics in the central and lower Bays during the spring and summer, and observe the response of the Bay's phytoplankton to river runoff, as well as provide a more thorough assessment of ODAS and a single set of algorithm coefficients.

One clearly identified problem in the Chesapeake Bay, similar to other impacted estuaries, is an elevated input of dissolved inorganic nutrients and a concomitant increase in the biomass of microscopic algae (Boynton et al. 1982; D'Elia et al. 1986; Harding et al. 1986; Correll 1987; Fisher et al. 1988). The phytoplanktonic algae accumulate to very high concentrations in spring as they use nutrients and light for to produce more algal biomass through photosynthesis. Part of the highly publicized cleanup effort in the Bay involves reduction of nitrogen and phosphorus inputs with the goal of reducing algal biomass by the turn of the century. The detection of responses of phytoplankton biomass densities to this effort is, therefore, an important element of the attempt to improve water quality and reclaim lost productivity.

Bay-wide Biomonitoring Programs have been developed by the states of Maryland and Virginia to track changes in physical, chemical and biological characteristics of the estuary that are expected to occur as results of the cleanup effort. Among the parameters that are monitored are chlorophyll a as a measure of algal biomass, and water transparency to determine the depth of the photic layer, the part of the water column in which active photosynthesis occurs. While these surveys generate data to provide twice monthly snapshots of conditions in the Chesapeake as a whole, they are restricted in time and space by expense and logistics. It is likely that ephemeral or localized events, such as algal blooms that may persist for only days to weeks, are missed or undersampled because of the Bay's size and complexity. To date, no affordable alternative to shipboard monitoring has been used to assess the size and longevity of these event-scale phenomena. The importance of phytoplankton blooms lies in the tremendous amount of organic carbon associated with them and in the link of this primary production to other processes, among them nutrient utilization, fisheries productivity, and the seasonal development of anoxic conditions.

A total of 16 flights covering the Bay were conducted between mid March and mid July 1989. Improved data processing procedures were developed using an image workstation. ODAS observations were compared with the extensive in-situ Bay monitoring data when they were coincident within two days and 0.01 degrees latitude and 0.005 degrees longitude (about 5 km by 7km). These wide limits were chosen to maximize the number of comparison points, and are much larger than ideal sea-truth comparison criteria. The results of the comparison are shown in Figure 2. The RMS deviation was found to be 0.243 log pigment concentration, or a factor of 1.7. Considering the broad limits place on temporal and spatial coincidence, widely diverse atmospheric and water optical conditions during the measurement period, this fit to a single set of algorithm coefficients is very pleasing. When sufficient contemporaneous aircraft and ship data permit the derivation of time and region specific coefficients, the fit is of course much tighter, as evidenced by figure 1.

Figure 3 is an example of pigment concentration patterns observed with the ODAS sensor produced using an objective interpolation between ODAS flight lines. The distributions clearly indicate higher spatial variability than could be sampled using only in-situ stations. The overall patterns of pigment distribution in 1989 were substantially different from recent annual patterns, a result of anomalous phasing of peak river flow bringing freshwater, nutrients, and sediment into the Bay. Also, a very tight relationship between surface chlorophyll concentrations and integrated chlorophyll content of the total water column, using digital bathymetry for the Bay, was another key finding of the demonstration. This relationship can be used to calculate the total phytoplankton biomass in the Bay, of direct importance to anoxia studies, using only remotely sensed surface concentrations.

An even larger number of monitoring flights (22) were conducted in 1990 jointly with VIMS personnel. Analysis and comparison of those data with in-situ Biomonitoring data still underway due to the longer time to reduce the in-situ data sets.

Operation of the sensor in a vigorous program such as the Chesapeake Bay Coastwatch Program revealed the need for several improvements which have been implemented in the data logging system, including faster updating of Loran data. The Beaver aircraft, very cost effective and capable of slow cruising speeds, proved to be an adequate platform. However, its single engine makes it a bit unsuitable for coastal ocean flights. Additionally, since the ODAS was mounted on the centerline within the fuselage, some difficulties were experienced with stray oil droplet deposit on the radiometer windows. Although no pronounced effect on radiometric data quality was apparent, it was virtually eliminated by raising the mounts so that the sensor was further away from the airstream.



Comparisons of surface temperature with the infrared radiometer indicate an accuracy of from 0.5 to 1.0 degree, which is consistent with errors in making such measurements due to variations in haze and humidity, and in comparing bulk water temperatures with skin brightness temperatures.

The NASA GSFC role in these and future uses of the ODAS is primarily one of assuring instrument capability and calibration, and consultation on data analysis. Investigators provide funding, a suitable aircraft, arrange for installation, mission logistics, and data analysis. The demand for cost effective ocean color sensors with modest computing requirements for data reduction is growing. The ODAS system was used to collect ocean color data during the CITE-3 missions off the East Coast of the US and Northeast South America. The data, collected from the NASA Electra, is being used to provide a measure of ocean biogenic trace gas sources for the atmospheric chemistry program. Data are being analyzed by Dr. P. Matrai at University of Miami. During summer of 1991, ODAS will be used in an introductory marine remote sensing course for undergraduates sponsored by the University of Maryland Sea Grant Office, in conjunction with the Bay Monitoring Program. Additionally, GSFC PI's plan to use the ODAS on occasion for their own ocean color research programs.

#### EXPENDABLE OPTICAL DRIFTING BUOY FOR SATELLITE OCEAN COLOR SENSOR VALIDATION

The objective of this GSFC Director's Discretionary Fund project is to develop and test a low-cost optical drifter for validation of satellite ocean color observations from the planned SeaWiFS and EOS missions, and to evaluate its reliability and the methodology for using such observations in sensor validation. Calibration and knowledge of temporal change of satellite visible radiometers to the accuracy and precision demanded for colorimetry of the ocean exceeds proven capability of pre-launch and on-board techniques. While absolute calibrations are improving, "vicarious" calibration procedures which include atmospheric correction algorithms and ocean bio-optical observations will remain essential. To be most useful for monitoring satellite sensor performance, these observations should be available from a global array within a short time period. There is no optical measurement data base for use with ocean color radiometry comparable to the international SST network. We hope that this DDF activity will help define required system design and performance characteristics, and elicit private sector interest in development of commercial units for the research application use.

Our approach is to build a prototype unit and test this in a tethered mode this winter, followed by deployment of up to four units in drift mode later this year. Designs have been developed which will return observations made at <30 min intervals during the day via the ARGOS system. The observations include a) incident surface irradiance ( $E_{0+}$ ) over 400 - 700 nm; b) upwelled radiance ( $L_u$  1m) at 460, 490, and 520 nm; and c) sea surface or buoy temperature. Only three channels were selected for the prototype because a small three channel radiometer was available from previous programs, and to live within transmission bandwidth constraints. Increasing the number of spectral channels to 6 or 8 is recommended for future units. The choice of the wavelengths for the bands is based on the ODAS and curvature algorithm experience; future systems should better match the spectral bands planned for future ocean color satellites. Based on relationships of spectral radiance and spectral attenuation,  $L_u$  1m, $\lambda$  can be propagated to the surface to compute spectral reflectance ( $R_\lambda$ , to estimate pigment concentration) and water-leaving radiance ( $L_w$ , $\lambda$ , for comparison with satellite or aircraft data). The stated objective in satellite sensor programs is to derive  $L_w$ , $\lambda$  to within  $\pm 5\%$  precision.

The drifter prototype package is about 1.2 m long by 15 cm diameter, and is compatible with air-deployment. Life expectancy is 4-6 months. Creative sampling, data processing and compression, and transmitter control are used to maximize information return. The prototype unit is currently undergoing full end to end test of the optics, data processing, and data transmission systems on the roof at GSFC.

The area of greatest risk scientifically is lack of knowledge of calibration and precision of the radiometer due to biofouling of the submerged optical windows. Organic tin compounds have been used to coat windows on optical moorings successfully, and test windows and prototype drifter housing with biofoulant coated windows are undergoing further testing by the Navy's David Taylor Research Center in Annapolis and San Diego. Growth must also be prevented everywhere to prevent changes in buoyancy.

Interrogating store and forward satellite data collection systems, planned by several groups, for testing in the next few years would be a very attractive alternate to ARGOS for future versions. Higher data volumes, reduced power consumption, and precise simultaneity with satellite overpasses are expected benefits.

## REFERENCES

- Boynton, W. R., W. M. Kemp & C. W. Keefe. 1982. A comparative analysis of nutrients and other factors influencing estuarine phytoplankton production. *In: Kennedy, V. S. (ed.) Estuarine Comparisons*, pp. 69-90. Academic Press, New York.
- Campbell, J. W. & W. E. Esaias. 1983. Basis for spectral curvature algorithms in remote sensing of chlorophyll. *Appl. Optics* 22: 1084-1093.
- Campbell, J. W. & W. E. Esaias. 1985. Spatial patterns in temperature and chlorophyll on Nantucket Shoals from airborne remote sensing data, May 7-9, 1981. *J. Mar. Res.* 43: 139-161.
- Campbell, J. W. & J. P. Thomas. 1981. Chesapeake Bay plume study - Superflux 1980. Proc. MTG. NASA, NMFS, NOAA, U. S. Dept. Commerce. Williamsburg, Virginia 21-23 Jan. 1981. NASA Conf. Publ. 2188 NOAA/NEMP III 81 ABCDFG 0042.
- Campbell, J. W., C. S. Yentsch & W. E. Esaias. 1986. Dynamics of phytoplankton patches on Nantucket shoals: An experiment involving aircraft, ships and buoys. *In: Bowman, M. J., C. M. Yentsch, W. T. Peterson (eds.) Tidal Mixing and Plankton Dynamics. Lecture Notes on Coastal and Estuarine Studies*, Vol. 17.
- Correll, D. W. 1987. Nutrients in Chesapeake Bay. *In: Majumdar, S. K., Hall, L. W., Austin, H. M. (eds.) Contaminant Problems and Management of Living Chesapeake Bay Resources*, pp. 298-320. Pennsylvania Acad. Sci.
- D'Elia, C. F., J. G. Sanders & W. R. Boynton. 1986. Nutrient enrichment studies in a coastal plain estuary: phytoplankton growth in large-scale, continuous cultures. *Can. J. Fish. Aquat. Sci.* 43: 397-406.
- Esaias, W. E. 1980. Remote sensing of oceanic phytoplankton: present capabilities and future goals. pp 321-337 *In: P.J. Falkowski, (ed.) Primary Productivity in the Sea*, Plenum, N.Y. 531 pp.
- Fisher, T. R., Jr., L. W. Harding, Jr., D. W. Stanley & L. G. Ward. 1988. Phytoplankton, nutrients, and turbidity in the Chesapeake, Delaware, and Hudson estuaries. *Est. Coast. Shelf Sci.* 27: 61-93.
- Grew, G. W. 1981. Real-time test of MOCS algorithm during Superflux 1980. NASA Publ. CP-2188, p. 301.
- Harding, L. W., Jr. 1988. Progress report on ODAS Demonstration Project in summer 1988. 18 October Mtg. STAC Chesapeake Bay Remote Sensing Work Group, Springfield, Virginia.
- Harding, L. W., Jr. and E. C. Itsweire 1990. Report on aircraft remote sensing of Chlorophyll in the Chesapeake Bay During 1989 Using NASA's Ocean Data Acquisition System (ODAS). The Chesapeake Bay Institute, The Johns Hopkins University. 25 June, 1990.
- Harding, L. W., Jr., B. W. Meeson & T. R. Fisher, Jr. 1986. Phytoplankton production in two East coast estuaries: photosynthesis-light functions and patterns of carbon assimilation in Chesapeake and Delaware Bays. *Est. Coast. Shelf Sci.* 23: 773-806.

- Hoge, F. E. & R. N. Swift. 1981. Airborne simultaneous spectroscopic detection of laser-induced Raman backscatter and fluorescence from chlorophyll and other naturally occurring pigments. *Appl. Optics* 20: 3197-3205.
- Hoge, F. E., R. N. Swift & J. K. Yungel. 1986. Active-passive airborne ocean color measurement. 2: Applications. *Appl. Optics* 25: 48-57.

### Figure Captions

**Figure 1.** Comparison of chlorophyll pigment signals from ODAS and the Airborne Oceanographic Lidar (AOL) off Wallops Island, VA. Chlorophyll pigment concentrations were estimated to range from 0.1 to 5 mg/cubic meter.

**Figure 2.** Linear regression of log surface pigment concentration collected by the Biomonitoring Program in 1989 on ODAS -log G. From Harding and Itsweire, 1990.

**Figure 3.** Interpolated map of surface chlorophyll concentration from ODAS track data. Dark areas were not sampled. The aircraft was excluded from the Patuxent Naval Air Station restriction zone in the central Chesapeake on this date. From Harding and Itsweire, 1990.

MARCH 28, 1988 PASS 6-5

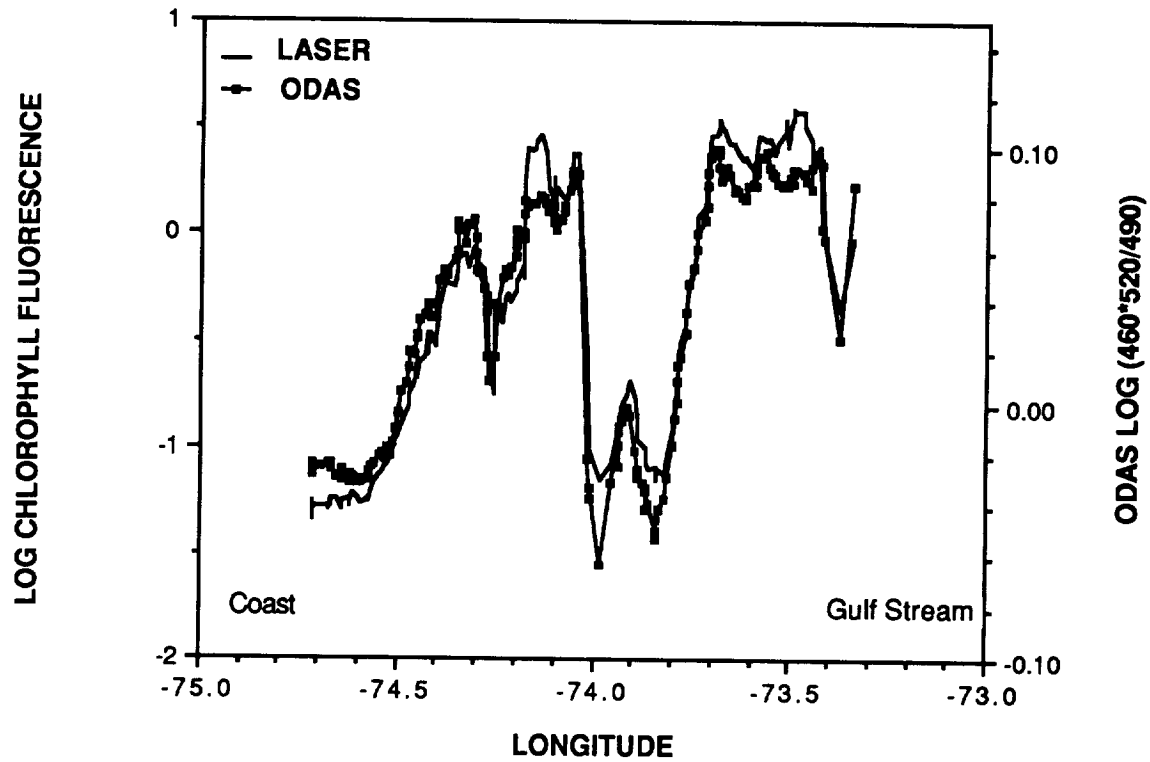


Figure 1. Comparison of chlorophyll pigment signals from ODAS and the Airborne Oceanographic Lidar (AOL) off Wallops Island, VA. Chlorophyll pigment concentrations were estimated to range from 0.1 to 5 mg/cubic meter.

1989 Biomonitoring Station Data  
Maryland and Virginia

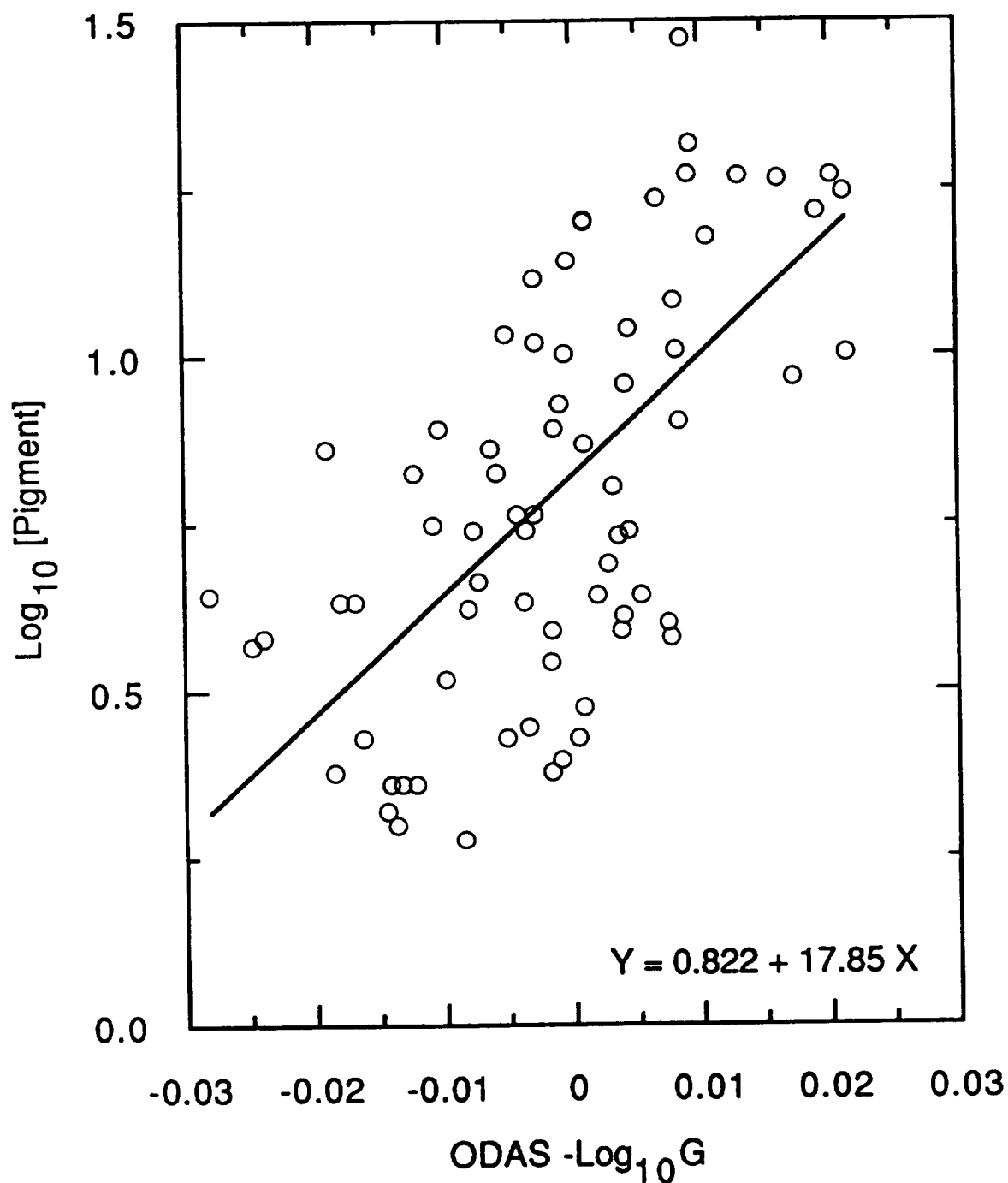


Figure 2. Linear regression of log surface pigment concentration collected by the Biomonitoring Program in 1989 on ODAS -log G. From Harding and Itsweire, 1990.

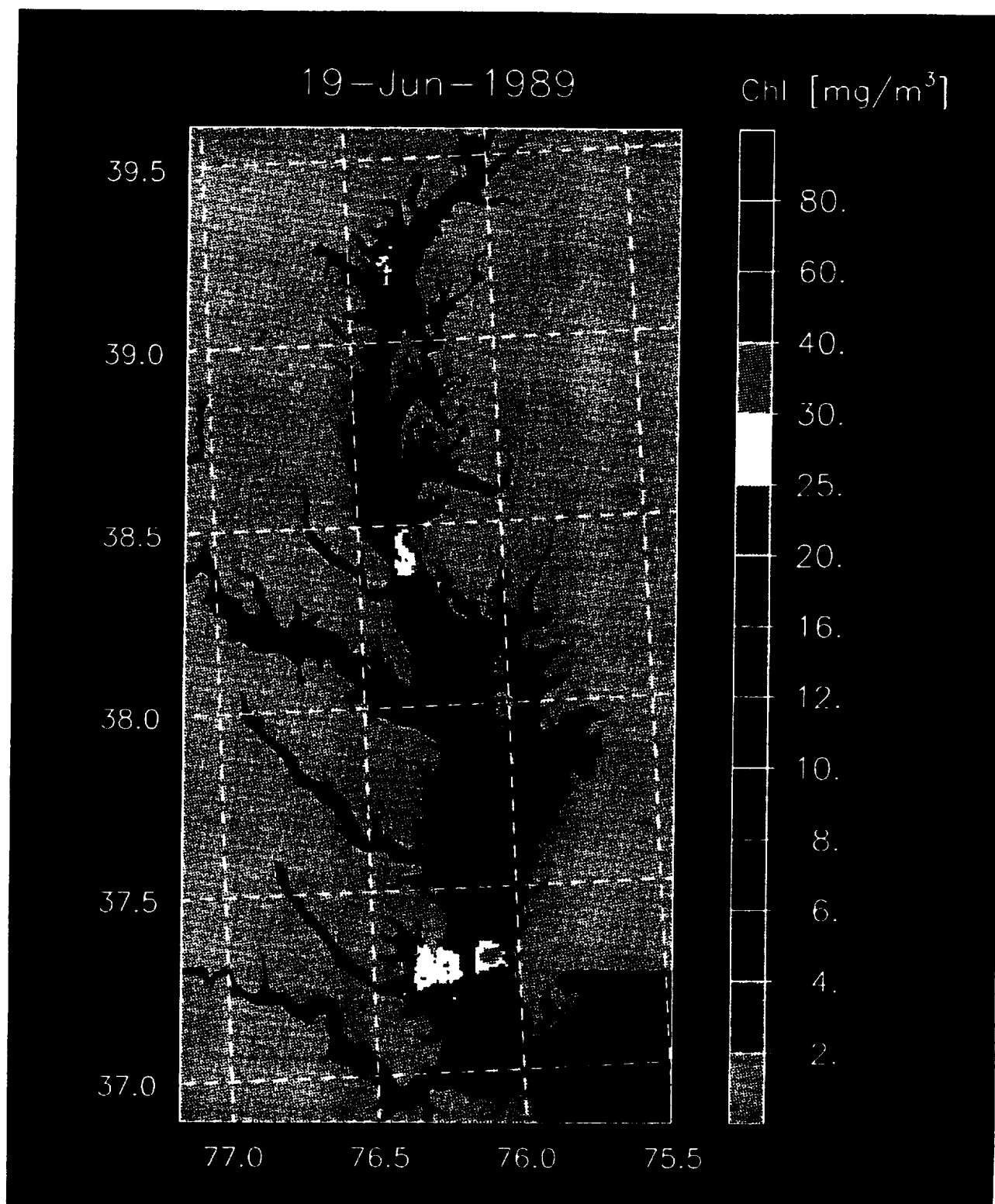


Figure 3. Interpolated map of surface chlorophyll concentration from ODAS track data. Dark areas were not sampled. The aircraft was excluded from the Patuxent Naval Air Station restriction zone in the central Chesapeake on this date. From Harding and Itsweire, 1990.



## MONITORING AND CONTROL OF ATMOSPHERE IN A CLOSED ENVIRONMENT

By R. Humphries and J. Perry

NASA, George C. Marshall Space Flight Center  
Thermal and Life Support Division

### INTRODUCTION

NASA is developing a manned orbiting space station for flight by the 21st century. Currently, the Space Station Freedom (S.S. Freedom), as it has been named, will be manned by an eight-person crew. An Earth-like atmosphere will be provided in 11 pressurized modules where the crew will live and work. The Environmental Control and Life Support System (ECLSS) is being developed to provide these surroundings artificially. The ECLSS is subdivided into six major subsystem groups to accomplish this task. These subsystem groups are composed of temperature and humidity control (THC), atmosphere control and supply (ACS), atmosphere revitalization (AR), water recovery and management (WRM), waste management (WM), and fire detection and suppression (FDS). As a whole, these subsystems supply, revitalize, condition, and monitor the respirable atmosphere; supply, recover, and condition water for hygiene and potable use; and collect, process, and store human waste for return to Earth.

Techniques have been adopted for some of these functions which are similar to those used in ground-based or earlier manned spacecraft applications. And, for the remainder, it has become necessary to develop new techniques. Those applications requiring new technologies for atmosphere monitoring and control in a closed environment will be discussed and their principal function onboard S.S. Freedom will be described.

### SYSTEM OVERVIEW

The closed environment inherent to spacecraft is conducive to contaminant buildup, especially over the long periods of time the S.S. Freedom is designed to be operational (~30 years). In such an environment, care must be taken to define all contaminant sources and sinks in detail. These sources range from human to equipment inputs. As part of the S.S. Freedom ECLSS activities, each of these sources is being characterized and analyzed in great depth via testing and analysis.

In striving to meet these challenges, the common denominators of all space system engineers are to minimize weight, power, and volume - three of the most valuable commodities of a spacecraft. Also, of great importance for S.S. Freedom are high reliability, minimization and ease of maintenance, and a low resupply penalty. With these in mind, designers are optimizing the S.S. Freedom ECLSS.

Undoubtedly, the greatest challenge the S.S. Freedom ECLSS designer faces is that of loop closure. For the first time in the history of the United States manned space flight program, both oxygen-and water-loop closure is planned. Oxygen-loop closure will be discussed because it is pertinent to the topic of air purity. Also, of special interest are atmospheric contaminant control and monitoring because of the new technologies necessary to efficiently accomplish these functions onboard S.S. Freedom. Although not addressed, other technologies are being developed which support these specific ECLSS functions. Examples of these other functions are electromechanical devices for new motor-driven valves with integral control devices, smart sensors, new 120-Vdc conversion devices, and others.

### OXYGEN LOOP CLOSURE

To minimize resupply demands, recycling is extremely important. Consequently, a string of equipment which can convert carbon dioxide to oxygen has been developed. Evolution of these technologies began in the 1960's and encompassed a number of different techniques. Competitive development of a

number of techniques, as shown in table 1, was funded until recently and brought the techniques to a predevelopment maturity for comparative testing. This equipment includes carbon dioxide removal and concentration, carbon dioxide reduction, and oxygen generation.

Low concentration carbon dioxide is extracted from the cabin air and concentrated by the removal subassembly and sent to a carbon dioxide reduction device. In this device, the carbon dioxide is catalytically reacted with hydrogen at high temperature, producing water vapor and a waste product. Water is condensed and ultimately fed in the form of hygiene water to a unit which decomposed it into its elemental parts of oxygen and hydrogen by an electrolytic process. The hydrogen is fed to the carbon dioxide reduction process for reaction with the carbon dioxide while the oxygen is returned to the cabin, thus closing the oxygen loop. Nitrogen purge streams for the carbon dioxide reduction and oxygen generation subassemblies are fed directly to the trace contaminant control subassembly for processing any residual hydrogen during startup and shutdown of these processes.

The water produced by the carbon dioxide reduction subassembly produces potable water for crew consumption. This water eventually is introduced into the air where it is removed as humidity condensate and reclaimed for crew consumption or into the waste management system. The water from the urine introduced into the waste system is reclaimed and processed for crew use during showering, hand washing, and other personal hygiene activities. Also, this water is used as a process feed stream to the oxygen generation subassembly, thus achieving loop closure.

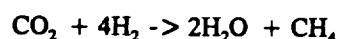
As a consequence of analytical studies combined with evaluation of comparative test results, baseline subassembly selections have recently been made. As indicated in table 1, the preferred approaches are four bed molecular sieve (4BMS) for carbon dioxide removal, the Sabatier reactor (Sa-CRS) for carbon dioxide reduction, and the static feed potassium hydroxide (KOH) electrolyzer (SFWES) for oxygen generation.

#### Carbon Dioxide Removal

The 4BMS, shown schematically in figure 1, uses a zeolite material (Linde 5A) which has been further modified by the supplier to produce a material which is extremely efficient in absorbing carbon dioxide. Unfortunately, this material also has a high affinity for water vapor. Consequently, the influent moist air must be dried. This is accomplished by using a two-media desiccant bed. This bed is made of silica gel and a second zeolite material (Linde 13X) with a high water removal efficiency. The influent air, carbon dioxide, and water vapor mixture is first passed through the desiccant bed where the dew point is reduced to a range of -30° F to -90° F. The carbon dioxide is then absorbed in the second 5A zeolite bed until the sorbent material becomes saturated. The carbon dioxide molecular sieve material is then desorbed by a combination pressure and temperature swing process which requires first evacuating the bed, then heating it to 400° F. The desorbed carbon dioxide is then purged from the bed in a concentrated form. The initial air passing through the bed after desorption is routed back through the desiccant bed to drive the absorbed moisture out of the desiccant bed and back into the cabin. During the desorption period of these two beds, the remaining beds are absorbing so that a continuous carbon dioxide removal capability exists at all times utilizing this four-bed concept.

#### Carbon Dioxide Reduction

Reduction of the carbon dioxide to water is achieved by a 950 °F Sabatier reactor process shown schematically by figure 2. The chemical reaction which takes place in this reactor is the following:

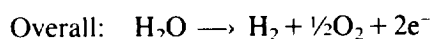
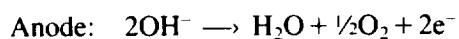
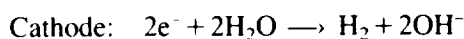


In this process, concentrated carbon dioxide at greater than 97-percent purity is fed into the reactor with hydrogen produced by the oxygen generator. A packed bed reactor containing a ruthenium on alumina catalyst provides the appropriate conditions for the process to proceed. This reaction forms water vapor and methane stoichiometrically. However, since the reaction is not 100 percent efficient, not all the carbon

dioxide is reacted and it exits with the water vapor and methane. A condenser at the reactor outlet condenses the water vapor to the liquid phase and it is pumped away to the potable water subsystem to be processed for crew consumption.

### Oxygen Generation

A KOH electrolyte-type generator shown schematically in figure 3 is used to separate water into hydrogen and oxygen. Electricity fed into the unit causes the electrolytic process to occur. At the same time, water enters a static water compartment where it evaporates through a Gortex membrane into the hydrogen cavity of the electrolysis cells. The chemical definition of how the water is separated into its elements by this process is the following:



By this process, gaseous hydrogen and oxygen is formed in respective compartments of the multicell unit. It is forced out of the unit by the pressure buildup caused by continuing reaction. The hydrogen produced from this process is fed to the Sa-CRS and the oxygen is supplied to the cabin for crew respiration.

### **ATMOSPHERIC CONTAMINATION**

Inherently, spacecraft must have closed environments. Since gas lost overboard must be produced at an expense of power, weight, and volume for closed loop processing equipment or through costly resupply from the ground, a tight environment is a must to reduce the loss and thus reduce resupply and resource requirements to make up for these losses. As a consequence, all contaminants evolved onboard a spacecraft from crew metabolism, material offgassing, and everyday housekeeping and work activities may build up over time. This buildup must be controlled and monitored to assure that the crew health is not compromised during their exposure to this environment of 180 days or more. This and a combination of other factors make contamination control and monitoring onboard S.S. Freedom a necessity. Of special concern is the large range of payloads and experiments planned over the S.S. Freedom life. This gives rise to the potential for a large variety of potential types as well as quantities of contaminants. Previous manned space flights by the United States have all been of short enough duration and/or in a large enough volume (i.e., Skylab) that metabolic methane buildup was of no concern. Complicating this for the S.S. Freedom is the fact that the new closed loop equipment generates contaminants such as hydrogen, methane, and carbon monoxide internally. All of which, if they leak, are not well removed by previously flown activated charcoal trace contaminant control media. As a consequence, for the first time a high temperature catalytic oxidation reactor or catalytic converter is being developed for the S.S. Freedom to be used in combination with activated carbon adsorption.

The need to monitor unexpected contaminants, long-term buildup of known contaminants, and contamination leakage gives rise to the need for an onboard real-time trace contaminant monitor. The need to monitor multiple atmospheric constituents points toward the desire to fly both a major constituent analyzer as well as a separate carbon monoxide monitor. The concern with long-term buildup of airborne particulates, where a clean room environment of less than or equal to 100,000 particles/ft<sup>3</sup> for particles between 0.5 and 100  $\mu$ m is desired in the station atmosphere, while hazardous aerosols may be generated by potential payloads such as gallium arsenide (GaAs) and mercury cadmium telluride (HgCdTe) crystal growth experiments, has stimulated both particulate and aerosol monitoring requirements on S.S. Freedom.

### Airborne Trace Contaminant Control

The trace contaminant control subassembly (TCCS) combines activated carbon adsorption with high temperature catalytic oxidation to achieve the most efficient long-duration airborne contamination control. This subassembly is shown schematically in figure 4. An activated charcoal impregnated with phosphoric acid is used to remove ammonia and high molar volume organic contaminants. Other contaminants which are not easily removed by the charcoal are removed in a 600<sup>0</sup> F to 700<sup>0</sup> F catalytic reactor. This reactor removes primarily methane, hydrogen, and low molar volume compounds while converting carbon monoxide to carbon dioxide. The reactor utilizes a palladium on alumina catalyst in a packed bed to achieve its function. Methane oxidation efficiencies as high as 95 percent per pass have been experienced using predevelopment and prototype testing. The reactor effluent is scrubbed by a lithium hydroxide (LiOH) packed bed to remove any acidic reaction products from the air stream before it is exhausted to the spacecraft cabin.

### Atmospheric Composition Monitoring

Atmospheric composition is monitored for trace contaminants, major constituents, carbon monoxide, and particulates on a continuous basis at multiple locations in the S.S. Freedom. Four separate instruments provide these functions since it is difficult for any one instrument to monitor the entire spacecraft atmospheric composition. Table 2 summarizes the S.S. Freedom atmospheric monitoring capabilities. Each instrument is described briefly.

### Trace Contaminant Monitor

The trace contaminant monitor (TCM) instrument utilizes a gas chromatography/mass spectrometry principle to analyze both species and quality for mass to charge ( $m/e$ ) ratio of 24 to 250. The TCM gas chromatograph uses two stages of sample preconcentration to detect low contaminant concentrations. The analysis is conducted in a cyclic, three-stage process. The sample is concentrated on a large sorbent trap and then further concentrated on a smaller sorbent trap. A small carrier volume is then used to transfer the sample to the carbonous fused silica gas chromatographic columns and then to a scanning double focusing magnetic sector mass spectrometer which scans a mass range from  $m/e$  24 to 250. The unit is capable of detecting at a resolution of 50 percent of the spacecraft maximum allowable concentration (SMAC) for each species monitored. Approximately 200 compounds are monitored through use of a software mass spectra fit library search. Any compounds which are not contained in the library are labeled as stranger compounds and may be analyzed offline if desired. The typical cycle time for this unit is 1 hour, and its software allows direct automated outputs of speciation and quantity. The samples are drawn through long, thin tubing from several locations within the S.S. Freedom cabin for analysis.

### Major Constituent Analyzer

The major constituent analyzer (MCA) is a mass spectrometer which utilizes a single focusing magnetic sector with a Faraday cup detector. It has a capability of monitoring in the mass to charge range of 1 to 48. This unit specifically focuses on oxygen, nitrogen, carbon dioxide, methane, hydrogen, and water vapor. The instrument's response time is less than 100 milliseconds (ms) for all gases except water which has a response time of 500 to 800 ms. Vacuum is maintained to the instrument by an ion pump which operates by surface absorption and chemical reaction of gases with the active metals of titanium and tantalum. It has a sampling rate of one sample per minute and draws samples from lines similar to those of the TCM.

### Carbon Monoxide Monitor

The carbon monoxide analyzer (COA) is a nondispersive infrared instrument using a dual isotope fluorescence technique. This instrument is used since carbon monoxide is not easily detected with mass spectrometry because it has approximately the same molecular weight as nitrogen. An air sample is introduced into the COA similarly to the TCM and MCA. In the instrument, an infrared source excites a

fluorescent cell which in turn produces radiation in the range for the most abundant isotope of carbon monoxide (carbon-12). The radiation passes through a chopper filter which contains the abundant and rare isotopes of carbon monoxide (carbon-12 and carbon-13) which results in the sample gas being exposed to alternating bursts of radiation which is absorbed by either the carbon-12 isotope or the carbon-13 isotope. Since carbon-13 is not very abundant, the radiation associated with it is not absorbed significantly by the sample and is used as a reference. The transmission and adsorption of the fluorescent radiation associated with the two isotopes is detected by a thermoelectrically cooled lead selenide (PbSe) detector. Since the degree of absorption is proportional to the carbon monoxide partial pressure, the detector signal is compared to calibration data to determine the partial pressure.

#### Particulate Counter Monitor

The particle counter monitor (PCM) is a compact commercial model based on light scattering. This device uses an AlGaAs laser diode to produce a light beam at a 780-nm (nanometer) wavelength which is scattered by particles in the air sample and detected by a photodetector. The light beam passes through several lenses and an aperture to produce a thin plane in the particle sensing zone. A light trap captures the main light beam after it passes the instrument's sensing zone. As each particle passes through the sensing zone and scatters light, the photodetector converts the scattered light energy to electrical pulses which have an amplitude corresponding to the particle size. The air is sampled through a dedicated line to continuously measure the total particle constituency from 0.5 to 100  $\mu\text{m}$ .

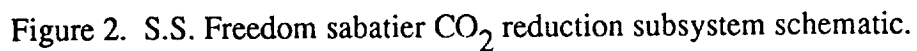
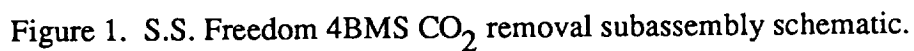
Table 1. O<sub>2</sub> closure competitive techniques

- CO<sub>2</sub> Removal
  - 4 Bed Molecular Sieve\*
  - Solid Amine Water Desorbed (SAWD)
  - Electrochemical Depolarized Cell (EDO)
- CO<sub>2</sub> Reduction
  - Sabatier/Advanced Carbon Reactor
  - Bosch Carbon Reactor
  - Sabatier\*
- O<sub>2</sub> Generation
  - Static Feed Water Electrolysis Subsystem (SFWES)\*
  - Anode Feed Solid Polymer Electrolyzer (AFSPE)

\* Baselined for SSF

Table 2. Atmospheric composition monitor (ACM) functional summary.

SUBASSEMBLY	INSTRUMENT	SPECIES MONITORED	RANGE/ ACCURACY	SAMPLE LOCATIONS	SAMPLE INTERVAL
Major Constituent Analyzer (MCA)	MS	N <sub>2</sub> , O <sub>2</sub> , H <sub>2</sub> O, H <sub>2</sub> , CH <sub>4</sub> , CO <sub>2</sub>	±5% of range	6	1 min/line
Trace Contaminant Monitor (TCM)	GC/MS	trace contaminants	0-50% of SMAC	6	30 min/line
Carbon Monoxide Analyzer (COA)	NDIR	CO	±1% of range	6	1 min/line
Particle Counter Monitor (PCM)	Light Scattering	particulates	0.5-100µm	1	continuous



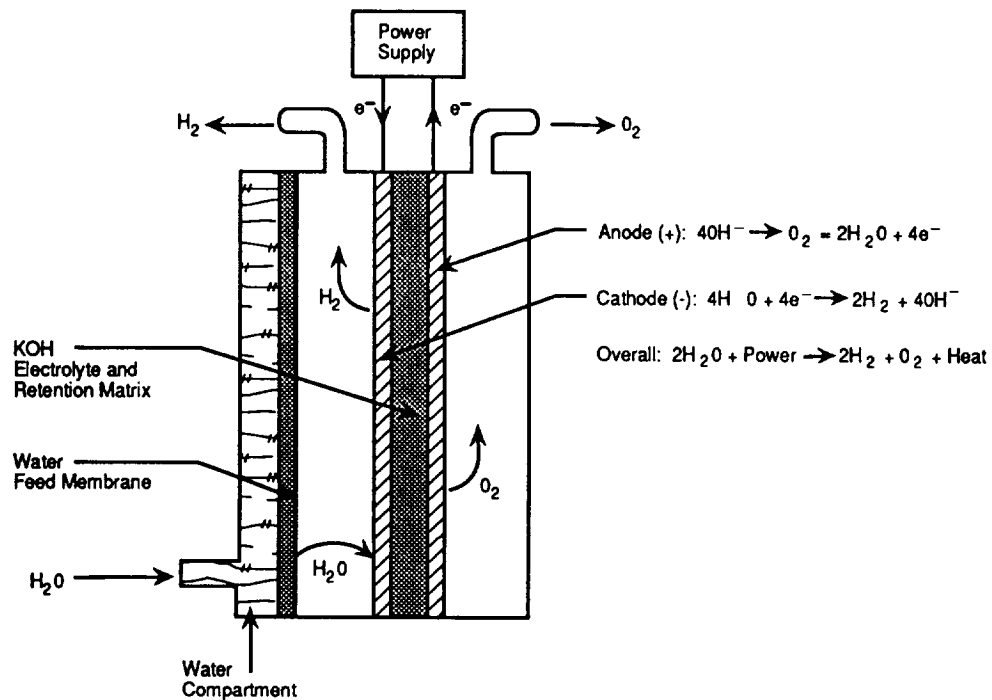


Figure 3. S.S. Freedom oxygen generator cross-sectional schematic.

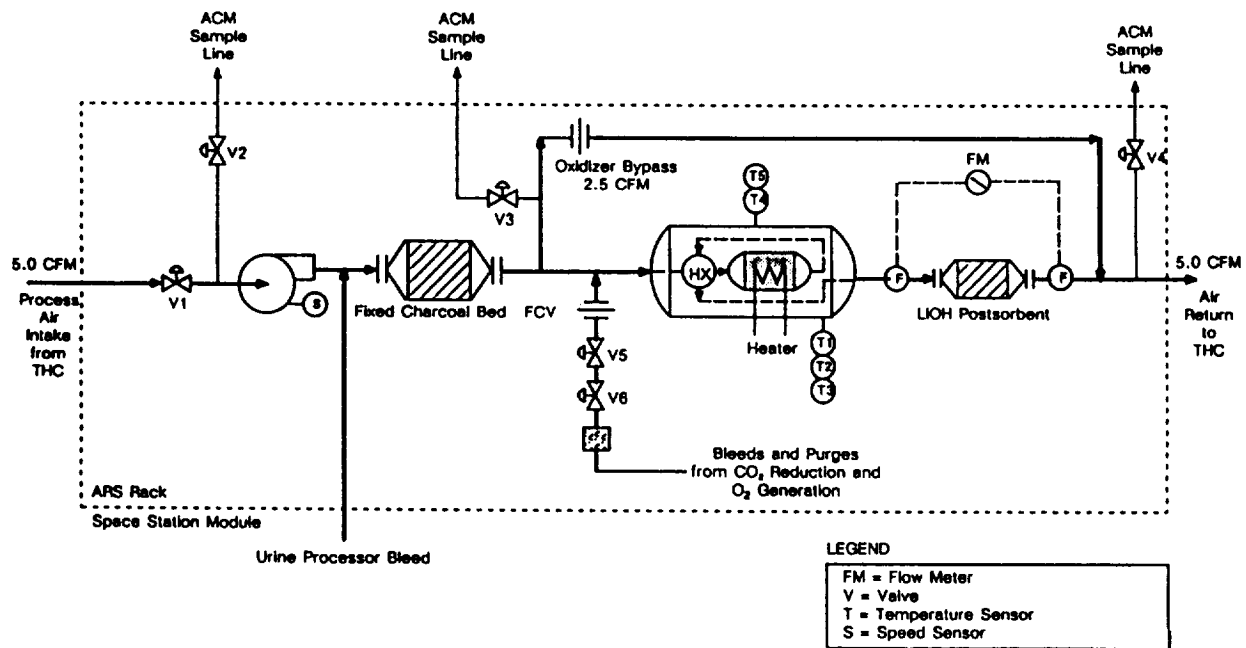


Figure 4. S.S. Freedom TCCS schematic.







## Report Documentation Page

1. Report No. NASA CP-3109, Vol. 1	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Technology 2000		5. Report Date March 1991	
		6. Performing Organization Code CU	
7. Author(s)		8. Performing Organization Report No.	
		10. Work Unit No.	
9. Performing Organization Name and Address  NASA Technology Utilization Division		11. Contract or Grant No.	
		13. Type of Report and Period Covered Conference Publication	
12. Sponsoring Agency Name and Address  National Aeronautics and Space Administration Washington, DC 20546		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract  Proceedings from the Technical Sessions of the Technology 2000 Conference and Exposition, November 27-28, 1990, Washington, DC. Volume 1 features 51 papers presented during eight concurrent sessions.			
17. Key Words (Suggested by Author(s))  technology transfer computer technology materials science robotics		18. Distribution Statement  Unclassified - Unlimited  Subject Category 99	
19. Security Classif. (of this report)  Unclassified	20. Security Classif. (of this page)  Unclassified	21. No. of pages  428	22. Price  A19







—

•